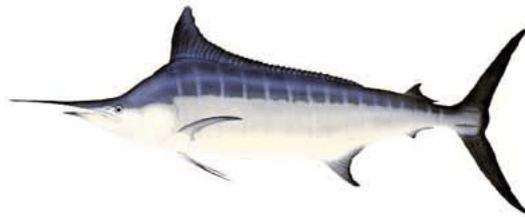


Preliminary Base-case Model in Stock Synthesis 3.30 for Consideration in the 2022 Western and Central North Pacific Striped Marlin (*Kajikia audax*) Stock Assessment using WCNPO Biological Parameters

Michelle Sculley*

*NOAA National Marine Fisheries Svc
Pacific Islands Fisheries Science Center

Email: michelle.sculley@noaa.gov



Abstract

A preliminary base-case model in Stock Synthesis 3.30 for Western and Central North Pacific (WCNPO) striped marlin (*Kajikia audax*) is described for consideration as the 2022 base-case model. The base-case model covers the Western and Central Pacific Fisheries Commission (WCPFC) management area north of the Equator from 1975 to 2020. It includes data from three International Scientific Committee for the Conservation of Tuna and Tuna-like Species (ISC) countries and other countries in aggregate from the WCPFC. This paper describes the data available for inclusion in the base-case model and the model using the biological parameters for the WCNPO stock. The model converges and appears to fit the data well. Initial diagnostics do not indicate major problems. Preliminary results suggest the WCNPO striped marlin stock is being fished above F_{MSY} and spawning stock biomass is below SSB_{MSY} .

Introduction

The International Scientific Committee for the Conservation of Tuna and Tuna-like Species (ISC) Billfish Working Group (BILLWG) has proposed to run a benchmark assessment on the Western and Central North Pacific striped marlin stock (*Kajikia audax*, MLS). Data were compiled from the International Scientific Committee for North Pacific Tuna and Tuna-like Species (ISC) member countries and other Western and Central Pacific Fisheries Commission (WCPFC) countries. Countries were asked to contribute catch, CPUE, and size-frequency data. It was decided to run the assessment using a single-sex, single-stock model in Stock Synthesis version 3.30 (Methot and Wetzel, 2013). Biological parameters were discussed by the billfish working group (BILLWG) at the data preparatory meeting in December 2021, where three growth and maturity curves representing estimates from the WCNPO, Eastern Pacific Ocean (EPO) and South West Pacific Ocean (SWPO) were discussed as significant uncertainty in stock structure and life-history parameters persist for Pacific stripe marlin. The WG agreed to develop base-case models for each regional life-history parameters to consider. The available data and the preliminary model results and diagnostics for the models using the WCNPO life history-parameters will be presented in this document for consideration at the ISC BILLWG MLS stock assessment meeting.

Methods

Spatiotemporal structure

The geographic area encompassed in the assessment for striped marlin was the Western and Central North Pacific Ocean bounded by the equator and the Western and Central Pacific Fisheries Commission management boundary at 150°W. Lacking conclusive evidence of a clearly defined stock boundary, the management unit with an eastern boundary of 150°W longitude was used as the definition of the stock for this assessment. Three types of data were used: fishery-specific catches, relative abundance indices, and length measurements. The fishery data were compiled for 1975-2020, noting that the catch data and length composition data were compiled and modeled on a quarterly basis. Several CPUE indices were also modeled as a quarterly index from the Japanese longline fleet. Available data, sources of data, and temporal coverage of the datasets used in the updated stock assessment are summarized in Figure 1. Further details are presented below.

Definition of fisheries

The fleet definitions for this assessment were generally unchanged from the 2019 MLS stock assessment in order to focus on exploring the uncertainty associated with other components of the model, with the exception of the division of the Japanese driftnet fleet into early/late and mid periods. A total of 25 fisheries that caught striped marlin were defined on the basis of country, gear type, location, and time period, where each fishery was assumed to target a distinct component of the stock. These fisheries included fourteen longline fisheries from Japan. Thirteen of these fleets are the results of the flexmix model applied to the Japanese offshore and distant-water longline data, which divided the data into areas and quarters based upon mean weight and CPUE. Nine quarter-area combinations were identified and two of these, Japan quarter 1 area 1 and quarter 3 area 1 were divided into the early and late periods. An additional longline fleet (JPNLL_Others) accounted for any other striped marlin longline catches. Three additional fleets from Japan included the driftnet catches in four fleets divided by quarter: quarters one and four and quarters two and three (JPNDF_Q14 and JPNDF_Q23) and time period early/late: 1975-1976 and 1994-2020, and mid: 1978-1993 and a fleet to encompass all other Japanese striped marlin catches (JPN_Others). There were also three fleets from Chinese Taipei: one for their distant water longline fleet (TWN_DWLL), one for their small-scale tuna longline fleet (TWN_STLL) and one other fleet for any additional catches (TWN_Others). There were two fleets from the United States: a single fleet for the Hawaii-based longline fleet (US_LL) and one other fleet (US_Others) which included handline and troll catches. Finally, there was one fleet for the various flags contained in the WCPFC management region not otherwise accounted for (WCPFC_Others). Descriptions and data sources to characterize the twenty-five fisheries that catch WCNPO striped marlin are also summarized in Table 1.

Catch

Catch was input into the model on a quarterly basis (i.e., by calendar year and quarter) from 1975 to 2020 for the 25 individual fisheries. Catch was reported in terms of catch biomass (mt) for all fisheries, with the exception of the Japanese offshore and distant water longline fleets (JPNLL F1-11) and the Japanese driftnet fleet from 1977 to 1993 (JPNDF Mid F23-24) for which catch was reported as numbers of fish caught.

Three countries (i.e., Japan, Chinese Taipei, and the USA) provided national catch data (Hirotaka Ijima, NRIFSF, personal communication; Yi-Jay Chang, NTU, personal

communication; Ito, PIFSC, personal communication). Striped marlin catches for all other fishing countries were collected from WCPFC category I and II data (Peter Williams, SPC, personal communication).

The resulting best available data on striped marlin catches by fishery from 1975-2020 were tabulated and are shown in Figure 2. The historical maximum and minimum annual striped marlin catches were 10,592 metric tons in 1975 and 2,217 metric tons in 2018, respectively. From 1975 to 1993, the Japanese driftnet fishery harvested approximately 40% of the total annual catch. Overall, annual catches of WCNPO striped marlin have generally declined since 1975. The annual catch of striped marlin in the WCNPO averaged about 2,500 metric tons in the period since the last assessment (2018-2020).

Relative Abundance Indices

Relative abundance indices for WCNPO striped marlin based on standardized CPUE were prepared for this assessment. A finite mixture model analysis was used to identify nine different area-quarter combinations based upon the weight and CPUE of striped marlin caught in the Japanese offshore and distant water longline fleets. Japanese CPUE data were standardized in two area-quarters (area one quarter one and area one quarter three) as well as pre- and post-1993 when Japanese logbook reporting requirements were changed (Ijima and Kanaiwa, 2019; Ijima and Koike, 2022).

Operational fishing data collected in the Hawaiian longline fishery by fishery observers in 1995-2020 were used for CPUE standardization of US longline fleets (Sculley, 2022). The fishery operates in two sectors; a shallow-set sector targeting swordfish and a deep-set sector targeting tunas. Striped marlin are caught as bycatch in both sectors. These data were standardized into a single CPUE time series including factors that accounted for much of the variability between sectors. The Chinese Taipei distant-water longline fleet was standardized from 1995-2020 using a Vector-Autoregressive Spatio-Temporal model (Lee et al., 2022b).

Visual inspection of all indices showed an overall decreasing trend with the last 5-10 years showing a relatively flat trend with the exception of the Chinese Taipei index which was highly variable with peaks in 2004 and 2013. Both of the early Japanese LL indices were relatively flat but variable through 1993.

The CVs for each CPUE index were assumed to be equal to the SE on the log scale. The minimum CV was scaled to a minimum of 0.2 or the square root of the residual variance (RSME) of what we would expect the assessment model to fit the CPUE index at best by adding a constant to each CV value. This was calculated as the square root of the residual variance of a loess smoother fit to each index (Francis 2011, Lee et al., 2014).

$$RSME_{smoother} = \sqrt{\left(\frac{1}{N}\right) \sum_{t=1}^N (Y_t - \hat{Y}_t)^2}$$

where Y_t is the observed CPUE in year t on the log scale, \hat{Y}_t is the predicted CPUE in year t from the smoother fit to the data on the log scale, and N is the number of CPUE observations. RSME values for each index are listed in **Error! Reference source not found.** If the input SE was greater than these values, it was left unchanged.

Size Composition Data

Quarterly fish length composition data from 1993–2020 for nine fisheries were available for the assessment and are summarized in Table 1. Length frequency data were compiled using 5-cm length bins from 50 to 230 cm. The lower boundary of each bin was used to define each bin for all composition data, and each observation consisted of the actual number of striped marlin measured. The new composition data were agreed upon at the BILLWG data workshop as the best available scientific information for the 2022 stock assessment.

Figure 3 shows the quarterly length compositions. Most of the fisheries caught small (<150cm) individuals. The aggregate length composition distributions were relatively consistent between fleets, with the exception of the US Longline fleet (Figure 4). Most longline size distributions had a single mode around 150-160cm. The US longline fleet was bimodal with peaks around 110cm and 140cm EFL. Data were fit using a multinomial error structure. Length composition data were weighted using the 2-stage process based upon the Francis (2011) method. In the first stage, the effective sample size was scaled to a mean of 25 by multiplying each number of samples by a constant. The second stage weighting was attempted based upon the T.A1.8 equation (Francis 2011) as calculated by the model using *r4ss*, an R package for plotting SS results (R version 4.0.5, R Core Team, 2021, *r4ss* version 1.42.0, Taylor et al., 2021).

Base-case model description

The assessment was conducted with Stock Synthesis (SS) version 3.30.18.00-SAFE

released 09/30/2021 using Otter Research ADMB 12.3 by Richard Methot (Methot and Wetzel, 2013). The model was set up as a single area and single-sex model with four seasons (quarters). Spawning was assumed to occur in May (month 5), while recruitment was assumed to occur in July (month 7). Age at recruitment was calculated based upon the model estimated average selectivity at age based upon the quarterly selectivity at length. The best-available biological parameters for the WCNPO stock were used with age-specific natural mortality (

Table 2. Mean input standard error (SE) in log-space (i.e., $\log(\text{SE})$) of lognormal error and root-mean-square-errors (RMSE), and additional variance added for the relative abundance indices for Western and Central North Pacific striped marlin used in the base-case model.

Fleet	RMSE	Mean input SE	Input+Additional Variance	Additional Variance	Fleet Name
26	0.157857	0.2	0.2	0	S01_JPNLL_Q1A1_Late
27	0.164315	0.2	0.2	0	S02_JPNLL_Q3A1_Late
28	0.202644	0.2	0.207998	0.007998	S03_US_LL
29	0.326102	0.2	0.305728	0.105728	S04_TWN_DWLL
30	0.072126	0.2	0.2	0	S05_JPNLL_Q1A1_Early
31	0.077925	0.2	0.2	0	S06_JPNLL_Q3A1_Early

Table 3 (Error! Reference source not found.) as agreed upon in the BILLWG Data Preparatory Meeting (ISC Report 2022). The maximum age of MLS was set to 15, the age at length L_1 was set to age 0.5, the CV of the growth curve was set to 0.14 for young fish and 0.08 for old fish, and the sex ratio at birth was assumed to be 1:1. The growth curve used a von Bertalanffy growth curve for ages 0.5-15 with a $K = 0.34$ and an $L_{inf} = 203$ cm EFL with the size at age 1 = 110 cm EFL. A Beverton-Holt spawner-recruit relationship was used with steepness (h) set at 0.87 and σ_R set at 0.6.

Thirty-one fleets were included in the model, 25 catch fleets and six survey fleets. Initial fishing mortality was estimated for F4. Main recruitment deviations were estimated from 1994-2020. The recruitment deviations were bias-adjusted based upon the estimates from Methot and Taylor (2011). Early recruitment deviations were estimated from 1960 to 1993 as the population was not at equilibrium prior to the start of the model.

The population model and the fishery length data had 37 five cm length bins from 50-230+ cm. The population had 16 annual ages from age 0 to 15. There were no age data. Fishery size data were used to estimate selectivity patterns, which controlled the size distribution of the fishery removals. Two different selectivity patterns were used based upon the best fit to the size composition data and CPUE indices. The Japanese driftnet fleets F13 and F14 and Chinese Taipei deepwater longline fleet F18 used an asymptotic logistic selectivity pattern. Using a more flexible double normal selectivity pattern resulted in the logistic shape, and therefore the simpler pattern was used for the fleets. All other fleets with size data were estimated as six-parameter double normal (dome-shaped) selectivity patterns. In addition, a cubic spline selectivity pattern with 3, 4, and 5 parameters was explored for the US Hawaii longline fleet, but these model runs failed to converge and so were discarded. Survey selectivity patterns mirrored their respective catch fleets (

Table 6. Comparison of $\ln(R_0)$ and B_0 estimates from various model runs using different combinations of input data. Converges? Indicates if a positive definite Hessian was estimated.

).

Model estimated time series of total biomass (B in metric tons, $mt = 1000 \text{ kg}$), age 1+ total biomass (B_{1+} mt), female spawning biomass (SSB mt), and recruitment (R in 1000s of fish) were tabulated on an annual basis. The annual exploitation rate was calculated as $Catch/B_{1+}$. Stock status indicators were calculated based upon MSY and not a target reference level.

Convergence Criteria and Diagnostics

The model was assumed to have converged if the standard error of the estimated parameters could be derived from the inverse of the negative Hessian matrix. Various convergence diagnostics were also evaluated. Excessive CVs ($>50\%$) on estimated parameters would suggest uncertainty in the parameter estimates or model structure. A gradient of >0.001 would suggest poorly fit parameter estimates. The correlation matrix was also evaluated to identify highly correlated ($>95\%$) and non-informative (<0.01) parameters. Parameter estimates hitting bounds of the prior was also indicative of poor model fit.

Several diagnostics were run to evaluate the fit of the model to the data. An Age-Structure Population Model (ASPM) was used to evaluate the influence of the length composition data on the population trends (Carvalho *et al.*, 2017). The ASPM was also used to explore how each CPUE index informed the population trends by running one-off ASPMs for each index. Profiling the likelihood on R_0 , where the R_0 is fixed at a range of values around the maximum likelihood estimate and then the likelihood is estimated, was used to identify influential data components (Lee *et al.*, 2014). A runs test was used to evaluate randomness in the residuals of the CPUE data (Carvalho *et al.*, 2021). Residual plots and plots of the observed vs expected data were examined to evaluate goodness-of-fit. Finally, a retrospective analysis and hindcast cross-validation were used to evaluate the predictive ability of the model (Carvalho *et al.*, 2021).

Results

Model fit

The WCNPO base-case model ran in about 10 minutes, estimated 102 parameters, and had a total likelihood of 1121.76. The inverse Hessian was positive definite, which allowed for the estimation of parameter standard deviations and suggests that the model converged, and the maximum gradient component was 0.0035, which is greater than 0.001. None of the parameter estimates hit a bound, no parameters had correlations above 0.95 and four selectivity parameters had correlations below 0.01. Thirty-two of thirty-four early recruitment deviations (1960-1993) and 14 of 27 of the main recruitment deviations had CVs > 50%. One of 31 selectivity parameters had CVs >50%. All of the parameters below the threshold for uncorrelated parameters also had CVs > 50%.

Fits to the abundance indices were relatively good, with no substantial divergences between the expected and estimated CPUEs (Figures 5 - 10). In addition all indices passed the runs test (Figure 1), which indicates that the residuals are likely random.

Estimated selectivity for each fleet are in Figures 12-Figure. Fits to the length composition data were also relatively good (Figures 16 - 20), although there are still problems fitting the US longline data (F16). The fit to the US size data is challenging because it is bimodal, however, current attempts to implement a cubic spline selectivity pattern have not been successful. Furthermore, the estimated mean size of fish caught in the Japanese driftnet fishery is slightly smaller than the observed data (Figure 4). However, all size composition time series passed the runs test (Figure 21).

Model estimates of age 1+ biomass show a slow decrease in biomass from 1975 to 1982, then biomass varied around MSY, declined to its lowest level in 1998, and has relatively stable since (Figure 22). Initial spawning stock biomass was estimated to be approximately 4,200 mt and virgin SSB was around 25,000 mt (Figure 23). Annual fishing mortality is reported as the average for fish ages 3-12 (Figure 24). Fishing mortality was above MSY for all except 6 years and excepting 2015, has been below F_{MSY} since 2014. Recruitment deviations suggested three periods of recruitment: high recruitment from 1975 to 1993, a period with little data and large variability around the estimates, average recruitment from 1994 to 2003 where recruitment varied around equilibrium recruitment, and low recruitment from 2004 to 2020. The log of the

deviations were generally between 0.6 and -0.6 (Figure 25). Current depletion, as estimated as the age 1+ biomass in 2020 compared to the virgin age 1+ biomass was estimated to be 0.09.

Diagnostics

Profiling on R_0 showed that the recruitment estimates were highly influential in the model results, and there was substantial conflict between the CPUE indices and the length composition data (Figures 26-28). The US data (CPUE and length comp) drive the model dynamics suggesting an $\ln(R_0)$ below 6.0, and Chinese Taipei data and Japanese size composition data suggesting an $\ln(R_0)$ around 7. Japanese CPUE data suggest an $\ln(R_0)$ around 6.1 (Tables 4-5).

Results from the ASPM model showed the same population trend as the full model during 1975-1993, which is the time period without any size composition data. After 1994, the ASPM biomass increases drastically and deviates completely from the base-case model (Figures 29 and 30). Further investigation running an ASPM with a single early and late index at a time indicates a similar pattern for all CPUE indices (Figure 31). The most likely explanation for this is that the population dynamics prior to 1994 are being driven by the two CPUE indices available (Figure 32). After 1994 catch decreases and CPUE for all fleets flattens out which would indicate that a stock is recovering. However, the size composition data indicate that the majority of the catch is juvenile fish. Continued removals of individuals before they have a chance to reproduce would continue to cause the stock to decline below MSY levels. This indicates that the size composition data are an integral component of the model, without which we would not have a full picture of the fishing effect on the stock.

The retrospective analysis indicates that a significant retrospective pattern exists for both biomass and fishing mortality (Mohn's $\rho = 0.2$ and -0.14 , respectively, Figure 33). Generally, biomass is overestimated and fishing mortality is underestimated. Results of the hindcast with cross-validation indicate that of the four CPUE indices at the end of the assessment horizon, only Chinese Taipei had reasonable predictive ability (MASE = 0.9), with all other fleets MASE > 1 (Figure 34). Comparing the predictive ability of the size composition data, two fleets had very good predictive ability (MASE < 0.5, F2 and F18), five had good predictive ability (MASE < 1 and > 0.5, F1, F4, F5, F6, F16) and one had poor predictive ability (MASE < 1, F14, Figure 35). F14 only had one datapoint in the analysis which likely explains why it had poor predictive ability.

The likelihood profile indicated that the US data components (CPUE and size composition data) are majority contributors to the likelihood. Attempts to downweight the US LL size composition data results in a model that fails to converge (due to a Hessian that is not positive definite). However, the ASPM models indicate that the US CPUE data suggest a similar trend as the other CPUE indices. To further explore the effect of the US data, models were run removing all US data except catch and fitting the parameters. While the models including the Japanese size data also failed to converge, the results for all the models indicated an estimated $\ln(R_0)$ between 6.2 and 6.4, and SSB_{zero} between 12500 and 15500 metric tons (Table 6). This suggests that even though the US data is a significant contributor to the likelihood, the estimated population size would be similar without the data.

One of the reasons that the model will converge with the US and Chinese Taipei data but not the Japanese size data, is because of the treatment of the fleets. When data is included as a quarter-area time series (i.e. Fleet 1 is quarter 1 area 1, fleet 2 is quarter 2 area two, etc) the model assumes that the fish caught in quarter 1 are the same fish caught in quarter 2 but with growth and movement. This means that to use quarter and specific fleets, the areas must be consistent between quarters (Figure 36, Ijima and Kanaiwa 2019). This is not the case for the Japanese deepwater longline fleets, where the areas change in each quarter and may not cover the same space. This makes it very difficult for SS to converge. In future assessments, this fleet structure will need to be adjusted.

Compared to 2019 base-case model, the 2022 base-case model spawning stock biomass estimates are larger and decline at the start of the assessment (Figure 37). However, the trend and SSB/SSB_{MSY} after 1993 are fairly similar. The main driver of the difference at the beginning of the assessment is that the initial fishing mortality is estimated in 2022 but fixed in 2019, and Japanese driftnet catches have been corrected to lower values in 2022. Fishing mortality is less variable in the 2022 model compared to the 2019 model and is generally lower in 1993 to 2020, with the last five years of the 2022 model below F_{MSY} (Figure 38). F_{MSY} was estimated to be approximately equal for both models (0.63 for the 2022 model and 0.61 for the 2019 model). Recruitment deviations are very similar between the two models except in the very beginning of the model, but this is a period with very high variability (Figure 39).

Conclusions

The 2022 base-case model is relatively consistent with the previous stock assessments. Despite declining catches, relatively flat CPUE indices, and fishing below MSY, the spawning stock biomass has not shown notable recovery. The model suggests that the stock has been below SSB_{MSY} since 1981 with the exception of 1982, 1983, and 1990. After a sharp decline from 1990 to 1997, a period when the model transitions from sparse data to data-rich, the SSB is relatively flat. Fishing mortality was highly variable until around 1997, and then has decreasing until the present. The model suggests that the stock is currently very likely overfished (probability > 99%) but that overfishing is likely not occurring (probability = 93%, Figure 40). Diagnostics indicate that the population trend is being driven by the length composition data which allows the model to take into account that the majority of the catch are juvenile MLS. In general, though there are still some problems with the diagnostics, most are identifiable and will not be able to be addressed in this current assessment.

References

- Carvalho, F., A. E. Punt, Y.-J. Chang, M. N. Maunder and K. R. Piner (2017). Can diagnostic tests help identify model misspecification in integrated stock assessments? *Fisheries Research* 192: 28-40.
- Carvalho, F., Winker, H., Courtney, D., Kapur, M., Kell, L., Cardinale, M. Schirripa, M., Kitakado, T., Yemane, D., Piner, K.R., Maunder, M.N., Taylor, I., Wetzel C.R., Doering, K. Johnson, K.F. and Methot, R.D. (2021). A cookbook for using model diagnostics in integrated stock assessments. Doi: 10.1016/j.fishres.2021.105959
- Francis, R. I. C. C. (2011). Data weighting in statistical fisheries stock assessment models. *Canadian Journal of Fisheries and Aquatic Sciences* 68(6): 1124-1138.
- González-Armas, R., Alexander, K-T., and Agustín, H-H. 2006. Evidence of billfish reproduction in the southern Gulf of California, Mexico. *Bulletin of Marine Science* 79, 3, 705-717.
- Humphreys Jr., R. and Brodziak, J., (2022). Revised analyses of the reproductive maturity of female striped marlin, *Kajikia audax*, in the central North Pacific off Hawaii. ISC/21/BILLWG-2/07.
- Ijima H., (2021). Update Japanese data set for striped marlin stock assessment in the Western and Central North Pacific Ocean. ISC/21/BILLWG-02/04.

- Ijima, H., and Kanaiwa, M. (2019). Size-dependent distribution of Pacific striped marlin (*Kajikia audax*): The analysis of Japanese longline fishery data using the finite mixture model. ISC/19/BILLWG-1/9.
- Ijima, H. and Koike, H. (2022). CPUE Standardization for Striped Marlin (*Kajikia audax*) using Spatio-Temporal Model using INLA. (ISC/21/BILLWG-02/01).
- ISC (2012). Stock assessment of striped marlin in the Western and Central North Pacific Ocean in 2011, Report of the Billfish Working Group Stock Assessment Workshop. July, Sapporo, Japan. ISC/SAR/MLS/2012.
- ISC BILLWG. (2022). Report of the Billfish Working Group Workshop. 13, 15-18 December 2021.
- Kell, L. T.; Mosqueira, I.; Grosjean, P.; Fromentin, J-M.; Garcia, D.; Hillary, R.; Jardim, E.; Mardle, S.; Pastoors, M.A.; Poos, J.J.; Scott, F.; and Scott, R.D. (2007). FLR: an open-source framework for the evaluation and development of management strategies. ICES Journal of Marine Science, 64 (4): 640-646.
Doi: [10.1093/icesjms/fsm012](https://doi.org/10.1093/icesjms/fsm012)
- Kopf, R.K., Davie, P.S., Bromhead, D., and Pepperell, J.G. (2011). Age and growth of striped marlin (*Kajikia audax*) in the Southwest Pacific Ocean. ICES Journal of Marine Science, 68(9), 1884-1895.
- Lee, K., Yi, C-H., Wang, W-J., Lu, C-Y., and Chang, Y-J. (2021a). Catch and size data of striped marlin (*Kajikia audax*) by the Taiwanese fisheries in the Western and Central North Pacific Ocean during 1958-2020. ISC/21/BILLWG-02/05.
- Lee, K., Hsu, J., Chang, Y-J. (2021b). CPUE standardization of stripe marlin caught by Taiwanese distant-water longline fishery in the Western and Central North Pacific Ocean during 1995 – 2020. ISC/21/BILLWG-02/02.
- Lee, H.-H., K. R. Piner, R. D. Methot Jr and M. N. Maunder (2014). Use of likelihood profiling over a global scaling parameter to structure the population dynamics model: An example using blue marlin in the Pacific Ocean. Fisheries Research 158: 138-146.
- Methot, R.D.; A'mar, T.; Wetzel, C.; and Taylor, I. (2017). Stock Synthesis User Manual Version 3.30.05-3.30.08. November 14, 2017.
- Methot, R. D. and I. G. Taylor (2011). Adjusting for bias due to variability of estimated recruitments in fishery assessment models. Canadian Journal of Fisheries and Aquatic Sciences 68(10): 1744-1760.

- Methot Jr, R. D. and C. R. Wetzel (2013). Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. *Fisheries Research* 142: 86-99.
- R Core Team 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Sculley M. (2022). Standardization of the Striped Marlin (*Kajikia audax*) Catch per Unit Effort Data Caught by the Hawaii-based Longline Fishery from 1994-2020 Using Generalized Linear Models ISC/21/BILLWG-02/03.
- Sevilla-Rodriguez. 2013. Ciclo reproductivo del marlin rayado baja california sur, Mexico. Master thesis.
- Sun, C.L., Hsu, W.S., Chang, Y.J., Yeh, S.Z., Chiang, W.C., and Su, N.J. (2011). Age and growth of striped marlin (*Kajikia audax*) in waters off Taiwan: A revision. Working paper submitted to the ISC Billfish Working Group Meeting, 24 May-1 June 2011, Taipei, Taiwan. ISC/11/BILLWG-2/07: 12p.
- Taylor, I.G.; Stewart, I.J.; Hicks, A.C.; Garrison, T.M.; Punt, A.E.; Wallace, J.R.; Wetzel, C.R.; Thorson, J.T.; Takeuchi, Y.; Ono, K.; Monnahan, C.C.; Stawitz, C.C.; A'mar, Z.T.; Whitten, A.T.; Johnson, K.F.; Emmet, R.L.; Anderson, S.C.; Lambert, G.I.; Stachura, M.M; Cooper, A.B.; Stephens, A.; and Klaer, N. (2021). r4ss package: R Code for Stock Synthesis. Version 1.28.0. URL <https://github.com/r4ss>

Tables and Figures

Table 1. Descriptions of fisheries catch and abundance indices included in the base case model for the stock assessment including fishing countries, time-period, and data sources.

Fleet No	Fleet name	Catch units	Size data	CPUE	Source
F1	F01_JPNLL_Q1A1_Late	N	Y	S01_JPNLL_Q1A1_Late	Ijima and Koike 2021 Ijima 2021a
F2	F02_JPNLL_Q1A2	N	Y	N	Ijima 2021a
F3	F03_JPNLL_Q1A3	N	N	N	Ijima 2021a
			(Mirror to F2)		
F4	F04_JPNLL_Q2A1	N	Y	N	Ijima 2021a
F5	F05_JPNLL_Q3A1_Late	N	Y	S02_JPNLL_Q3A1_Late	Ijima and Koike 2021 Ijima 2021a
F6	F06_JPNLL_Q4A1	N	Y	N	Ijima 2021a
F7	F07_JPNLL_Q1A4	N	N	N	Ijima 2021a
			(Mirror to F2)		
F8	F08_JPNLL_Q2A2	N	N	N	Ijima 2021a
			(Mirror to F4)		
F9	F09_JPNLL_Q3A2	N	N	N	Ijima 2021a
			(Mirror to F5)		
F10	F10_JPNLL_Q4A2	N	N	N	Ijima 2021a
			(Mirror to F6)		
F11	F11_JPNLL_Q4A3	N	N	N	Ijima 2021a
			(Mirror to F6)		
F12	F12_JPNLL_Others	B	N	N	Ijima 2021a
			(Mirror to F4)		
F13	F13_JPNDF_Q14_EarlyLate	B	Y	N	Ijima 2021a
F14	F14_JPNDF_Q23_EarlyLate	B	Y	N	Ijima 2021a
F15	F15_JPN_Others	B	N	N	Ijima 2021a
			(Mirror to F4)		
F16	F16_US_LL	B	Y	S03_US_LL	Sculley 2021 Russ Ito, pers. comm.
F17	F17_US_Others	B	N	N	Russ Ito, pers. comm.
			(Mirror to F16)		
F18	F18_TWN_DWLL	B	Y	S04_TWN_DWLL	Lee et al., 2021a

					Lee et al., 2021b
F19	F19_TWN_STLL	B	N	N	Lee et al., 2021a
			(Mirror to F18)		
F20	F20_TWN_Others	B	N	N	Lee et al., 2021a
			(Mirror to F14)		
F21	F21_WCPFC_Others	B	N	N	WCPFC yearbook
			(Mirror to F12)		
F22	F22_JPNLL_Q1A1_Early	N	N	S06_JPNLL_Q1A1_Early	Ijima and Koike 2021
			(Mirror to F1)		Ijima 2021a
F23	F23_JPNLL_Q3A1_Early	N	N	S07_JPNLL_Q3A1_Early	Ijima and Koike 2021
			(Mirror to F5)		Ijima 2021a
F24	F13_JPNDF_Q14_Mid	N	N	N	Ijima 2021a
			(Mirror to F13)		
F25	F14_JPNDF_Q23_Mid	N	N	N	Ijima 2021a
			(Mirror to F14)		

Table 2. Mean input standard error (SE) in log-space (i.e., $\log(\text{SE})$) of lognormal error and root-mean-square-errors (RMSE), and additional variance added for the relative abundance indices for Western and Central North Pacific striped marlin used in the base-case model.

Fleet	RMSE	Mean input SE	Input+Additional Variance	Additional Variance	Fleet Name
26	0.157857	0.2	0.2	0	S01_JPNLL_Q1A1_Late
27	0.164315	0.2	0.2	0	S02_JPNLL_Q3A1_Late
28	0.202644	0.2	0.207998	0.007998	S03_US_LL
29	0.326102	0.2	0.305728	0.105728	S04_TWN_DWLL
30	0.072126	0.2	0.2	0	S05_JPNLL_Q1A1_Early
31	0.077925	0.2	0.2	0	S06_JPNLL_Q3A1_Early

Table 3. Key life history, recruitment, and selectivity parameters for the WCNPO striped marlin models, the biological parameters used for this model are bolded. From Table 2 in the ISC BILLWG Data Preparatory report (2022).

Parameter	WCNPO (Revised)	SWPO	EPO	Reference
Growth_Age_for_L1	0.5	0.5	0.5	Refit Ijima (2021b)
Growth_Age_for_L2	15	15	15	Refit Ijima (2021b)
NatM	0.54 (0)	0.54 (0)	0.54 (0)	
	0.47 (1)	0.47 (1)	0.47 (1)	
	0.43 (2)	0.43 (2)	0.43 (2)	
	0.4 (3)	0.4 (3)	0.4 (3)	
	0.38 (4+)	0.38 (4+)	0.38 (4+)	
L_at_Amin_Fem_GP_1	110	115	74	Refit Ijima (2021b)
L_at_Amax_Fem_GP_1	203	212	184	Refit Ijima (2021b)
VonBert_K_Fem_GP_1	0.34	0.64	0.23	Refit Ijima (2021b)
CV_young_Fem_GP_1	0.14	0.14	0.14	ISC 2012
CV_old_Fem_GP_1	0.08	0.08	0.08	ISC 2012
Wtlen_1_Fem	4.68e-06	4.68e-06	4.68e-06	Sun et al. (2011)
Wtlen_2_Fem	3.16	3.16	3.16	Sun et al. (2011)
Mat50%_Fem	152.2¹	178.4 ²	181 ³ 166.5 ⁴	¹ Humphreys and Brodziak (2022) ² Kopf et al. (2012) ³ Gonzalez-Armas et al. (2006) ⁴ Sevilla-Rodriguez (MS 2013)
Mat_slope_Fem	-0.204	-0.204	-0.204	Humphreys and Brodziak (2022)
Fecundity	Proportional to spawning biomass	Proportional to spawning biomass	Proportional to spawning biomass	-
Spawning season	July	July	July	ISC 2012
R0	-	-	-	Estimate
Steepness	0.87	0.87	0.87	Brodziak et al. (2015)

Table 4. Relative negative log-likelihoods of abundance index data components in the base case model over a range of fixed levels of virgin recruitment in log-scale ($\log(R_0)$). Likelihoods are relative to the minimum negative log-likelihood (best-fit) for each respective data component. Colors indicate relative likelihood (green: low negative log-likelihood, better-fit; red: high negative log-likelihood, poorer-fit). Maximum likelihood estimate of $\log(R_0)$ was 6.31. See Table 1 for a description of the abundance indices.

$\ln(R_0)$	S01	S02	S03	S04	S05	S06
5	0.35	0.44	0	0.02	0.63	0.50
5.1	0.31	0.38	0.001	0.01	0.48	0.32
5.2	0.26	0.32	0.01	0.01	0.38	0.19
5.3	0.23	0.27	0.02	0.01	0.30	0.10
5.4	0.22	0.25	0.04	0	0.28	0.09
5.5	0.15	0.17	0.04	0.03	0.23	0.02
5.6	0.11	0.12	0.06	0.04	0.23	0.001
5.7	0.07	0.08	0.08	0.05	0.25	0
5.8	0.04	0.04	0.11	0.06	0.23	0.02
5.9	0.02	0.02	0.13	0.07	0.21	0.04
6	0.01	0.01	0.16	0.07	0.20	0.07
6.1	0	0	0.19	0.08	0.18	0.10
6.2	0.001	0.003	0.22	0.08	0.16	0.14
6.3	0.01	0.02	0.26	0.07	0.14	0.18
6.31	0.01	0.02	0.26	0.07	0.14	0.18
6.4	0.03	0.04	0.30	0.07	0.13	0.21
6.5	0.05	0.07	0.34	0.06	0.11	0.24
6.6	0.08	0.11	0.39	0.04	0.09	0.27
6.7	0.12	0.16	0.44	0.02	0.07	0.29
6.8	0.18	0.20	0.51	0.04	0	0.47
6.9	0.18	0.23	0.54	0.04	0.11	0.31
7	0.16	0.21	0.56	0.06	0.19	0.60

Table 5. Relative negative log-likelihoods of length composition data components in the base case model over a range of fixed levels of virgin recruitment in log-scale ($\log(R_0)$). Likelihoods are relative to the minimum negative log-likelihood (best-fit) for each respective data component. Colors indicate relative likelihood (green: low negative log-likelihood, better-fit; red: high negative log-likelihood, poorer-fit). Maximum likelihood estimate of $\log(R_0)$ was 6.31. See Table 1 for a description of the composition data.

ln(R0)	F01	F02	F04	F05	F06	F13	F14	F16	F18
5	0.90	0.52	1.67	0	0	0.73	0.53	1.68	0.38
5.1	0.90	0.57	1.63	0.03	0.05	0.82	0.70	1.38	0.36
5.2	0.88	0.62	1.58	0.07	0.10	0.90	0.85	1.14	0.35
5.3	0.87	0.66	1.54	0.11	0.14	0.97	1.00	0.94	0.33
5.4	0.88	0.64	1.42	0.16	0.14	0.89	1.02	1.36	0.29
5.5	0.81	0.74	1.46	0.18	0.23	1.13	1.31	0.53	0.30
5.6	0.77	0.80	1.44	0.22	0.28	1.23	1.49	0.24	0.29
5.7	0.74	0.84	1.41	0.25	0.32	1.33	1.65	0	0.28
5.8	0.72	0.87	1.37	0.29	0.35	1.36	1.73	0.01	0.27
5.9	0.70	0.88	1.31	0.32	0.36	1.37	1.76	0.14	0.25
6	0.67	0.87	1.24	0.36	0.37	1.35	1.76	0.38	0.24
6.1	0.63	0.84	1.16	0.39	0.37	1.31	1.71	0.74	0.22
6.2	0.59	0.80	1.07	0.42	0.36	1.23	1.62	1.25	0.20
6.3	0.53	0.73	0.95	0.45	0.34	1.11	1.48	1.90	0.18
6.31	0.53	0.73	0.95	0.45	0.34	1.11	1.47	1.94	0.18
6.4	0.47	0.64	0.82	0.47	0.31	0.97	1.29	2.71	0.16
6.5	0.40	0.53	0.67	0.50	0.26	0.79	1.05	3.70	0.13
6.6	0.33	0.40	0.51	0.52	0.21	0.58	0.77	4.86	0.10
6.7	0.25	0.25	0.32	0.54	0.15	0.35	0.45	6.20	0.06
6.8	0.06	0.12	0.18	0.59	0.09	0.16	0.18	7.35	0.03
6.9	0.13	0.03	0	0.58	0.03	0	0	8.44	0
7	0	0	0.03	0.55	0.02	0.10	0.12	8.04	0.01

Table 6. Comparison of $\ln(R0)$ and $B0$ estimates from various model runs using different combinations of input data. Converges? Indicates if a positive definite Hessian was estimated.

Model	Ln(R0)	Bzero	Converges?
Base - all data, no weighting	6.29	15565	Y
No US Size	6.48	14992	N
Only TWN data	6.41	14010	Y
TWN plus JPN index 1,2, JPN size 1	6.44	14425	Y
JPN and TWN CPUE, Only TWN Size	6.51	14999	N
JPN and TWN CPUE, Only JPN Size	6.34	15565	N
Base - all data, weighted	6.31	12573	Y
All Data, US cubic spline (5 params)	6.36	13301	N

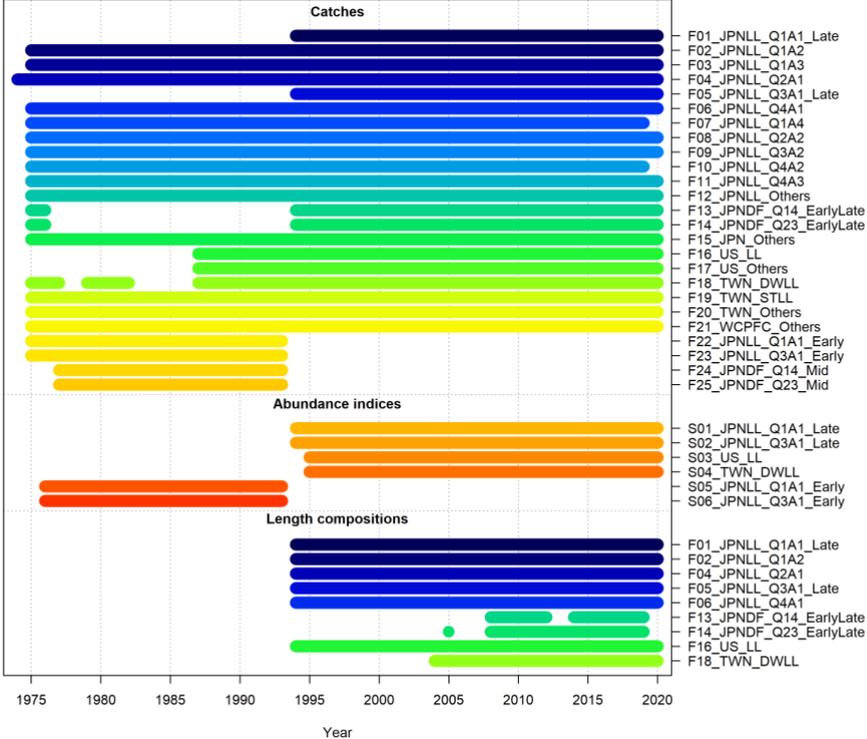


Figure 1. Catch, CPUE index, and size composition data included in the 2022 WCNPO striped marlin stock assessment.

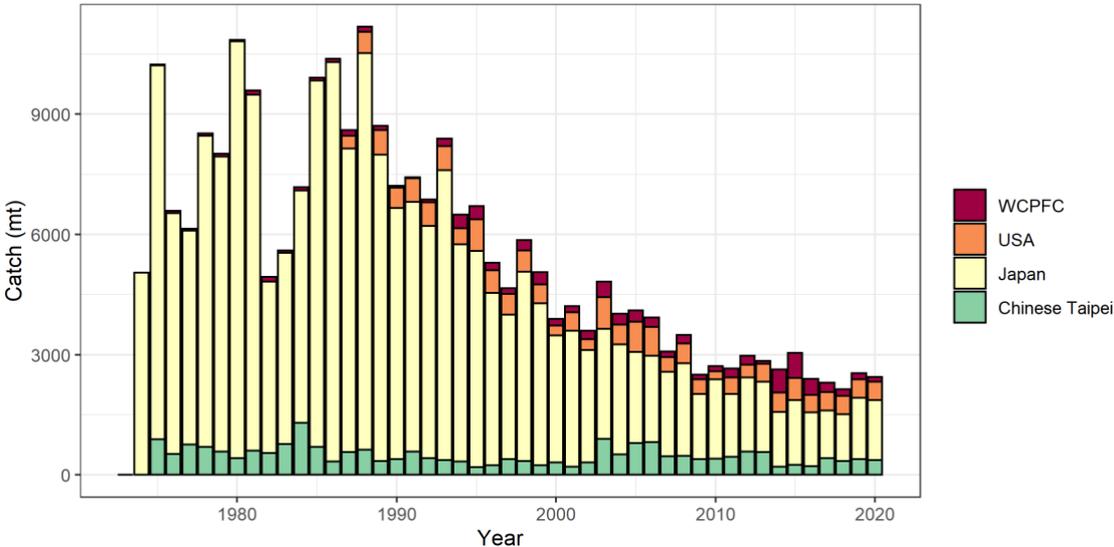


Figure 2. Annual catch of WCNPO striped marlin by country or RFMO and gear used in the 2022 base-case assessment model.

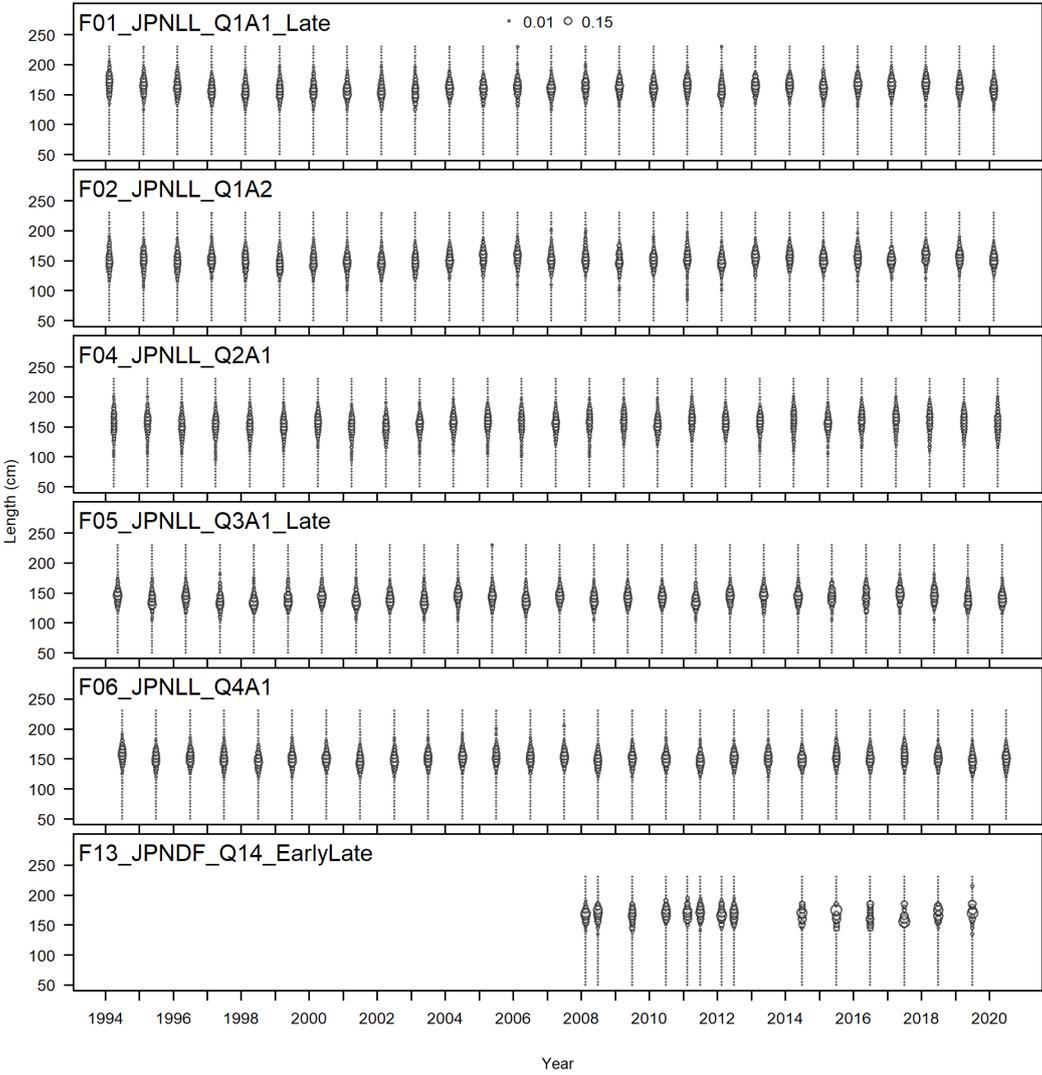


Figure 3. Length Composition data available in 5cm size bins for the 2022 WCNPO striped marlin stock assessment.

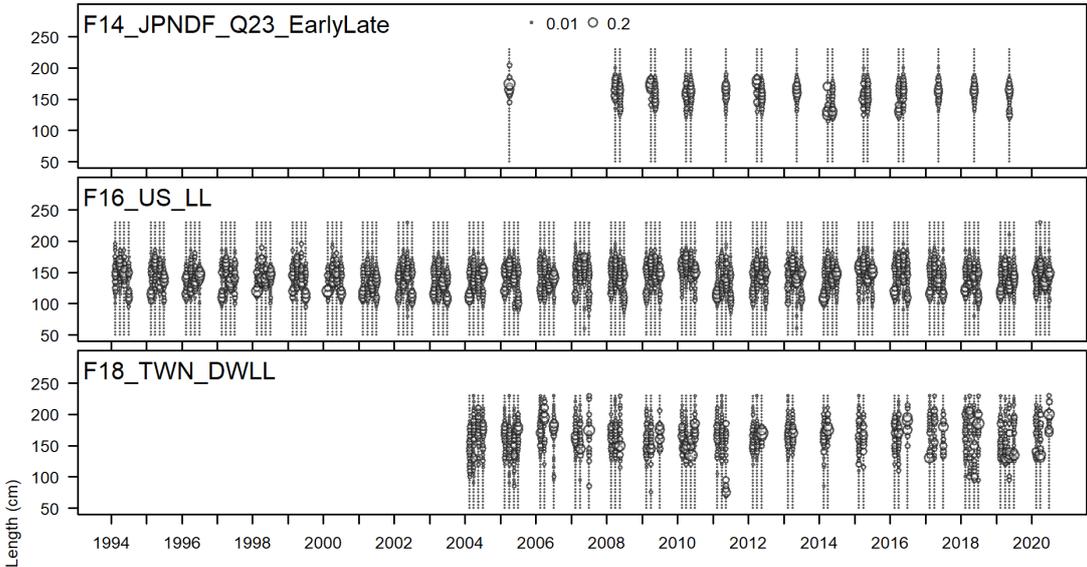


Figure 3. Cont.

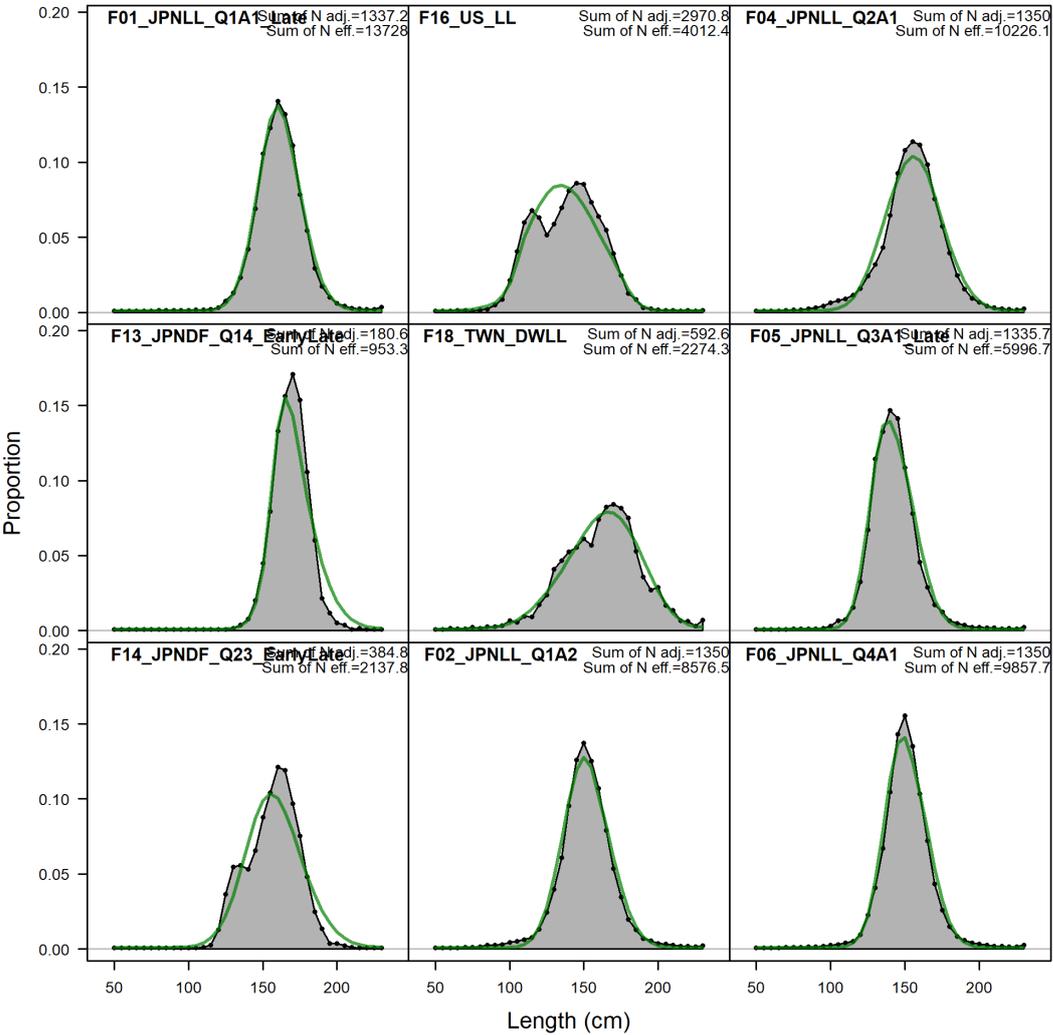


Figure 4. Aggregate length composition data available for the 2022 WCNPO striped marlin assessment, grey shading indicates observed data, green line indicates expected distribution based upon the estimated selectivity.

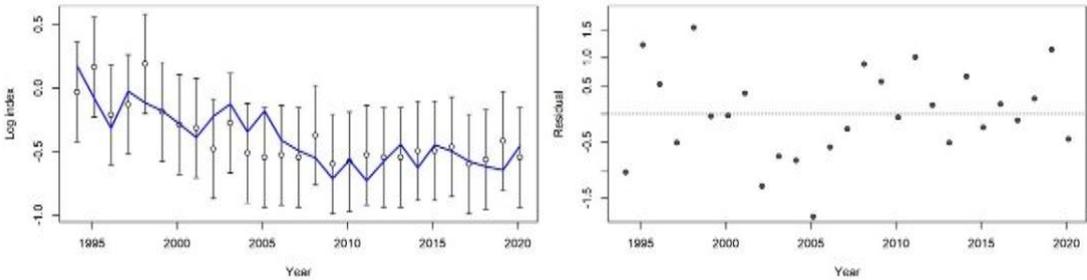


Figure 5. Fit to the S1 Japanese Late Q1A1 LL CPUE index. Left is the input CPUE with CV and the model fit CPUE (blue line). Right is the annual residuals of that fit.

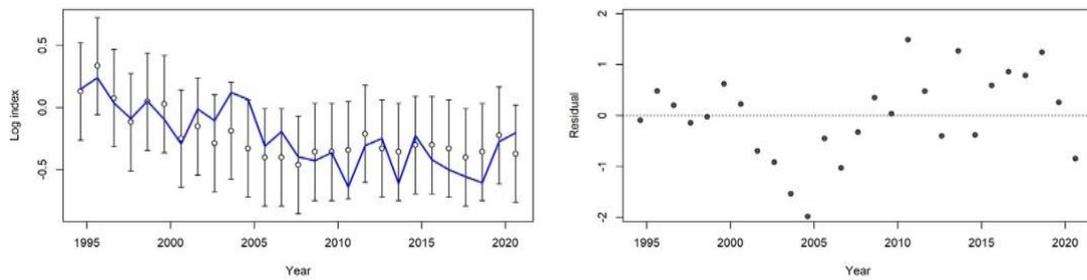


Figure 6. Fit to the S2 Japanese Late Q3A1 LL CPUE index. Left is the input CPUE with CV and the model fit CPUE (blue line). Right is the annual residuals of that fit.

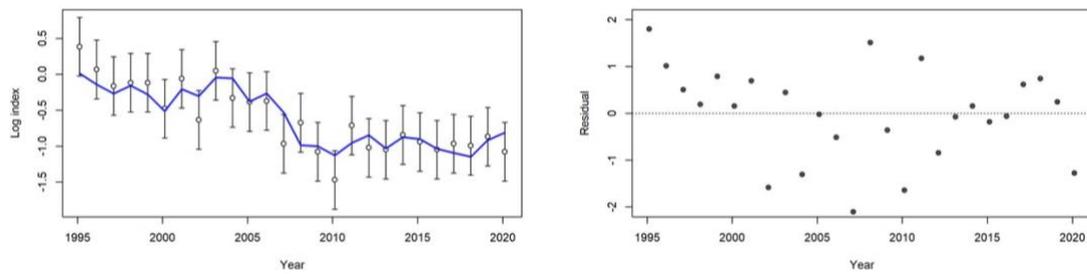


Figure 7. Fit to the S3 US Hawaii LL CPUE index. Left is the input CPUE with CV and the model fit CPUE (blue line). Right is the annual residuals of that fit.

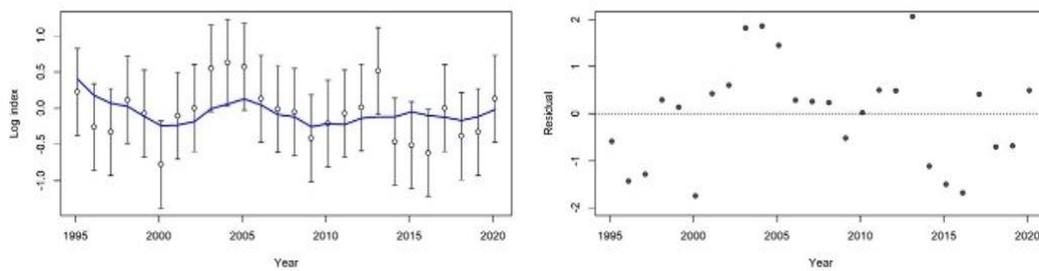


Figure 8. Fit to the S4 Chinese Taipei DWLL CPUE index. Left is the input CPUE with CV and the model fit CPUE (blue line). Right is the annual residuals of that fit.

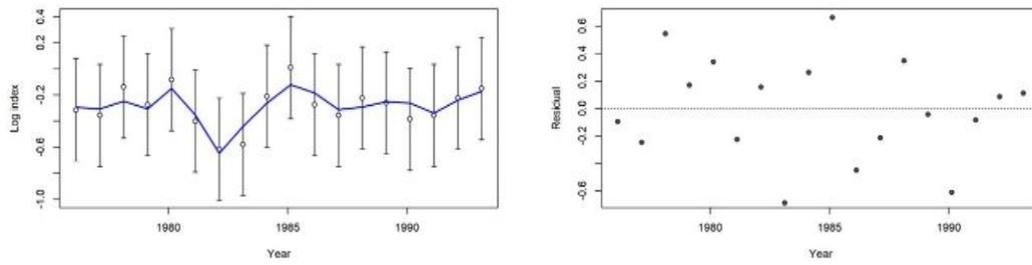


Figure 9. Fit to the S5 Japanese Early Q1A1 LL CPUE index. Left is the input CPUE with CV and the model fit CPUE (blue line). Right is the annual residuals of that fit.

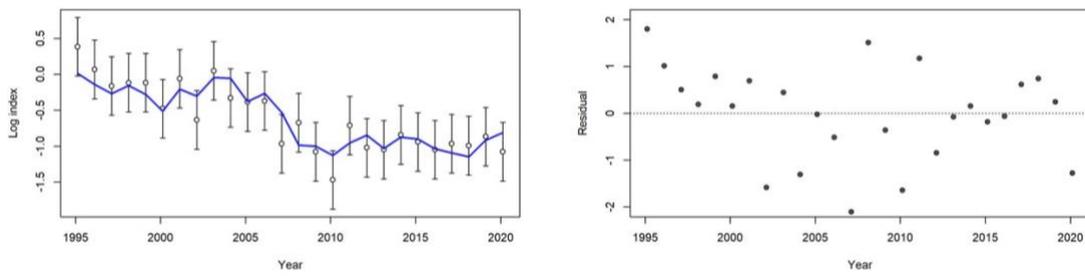


Figure 10. Fit to the S6 Japanese Early Q3A1 LL CPUE index. Left is the input CPUE with CV and the model fit CPUE (blue line). Right is the annual residuals of that fit.

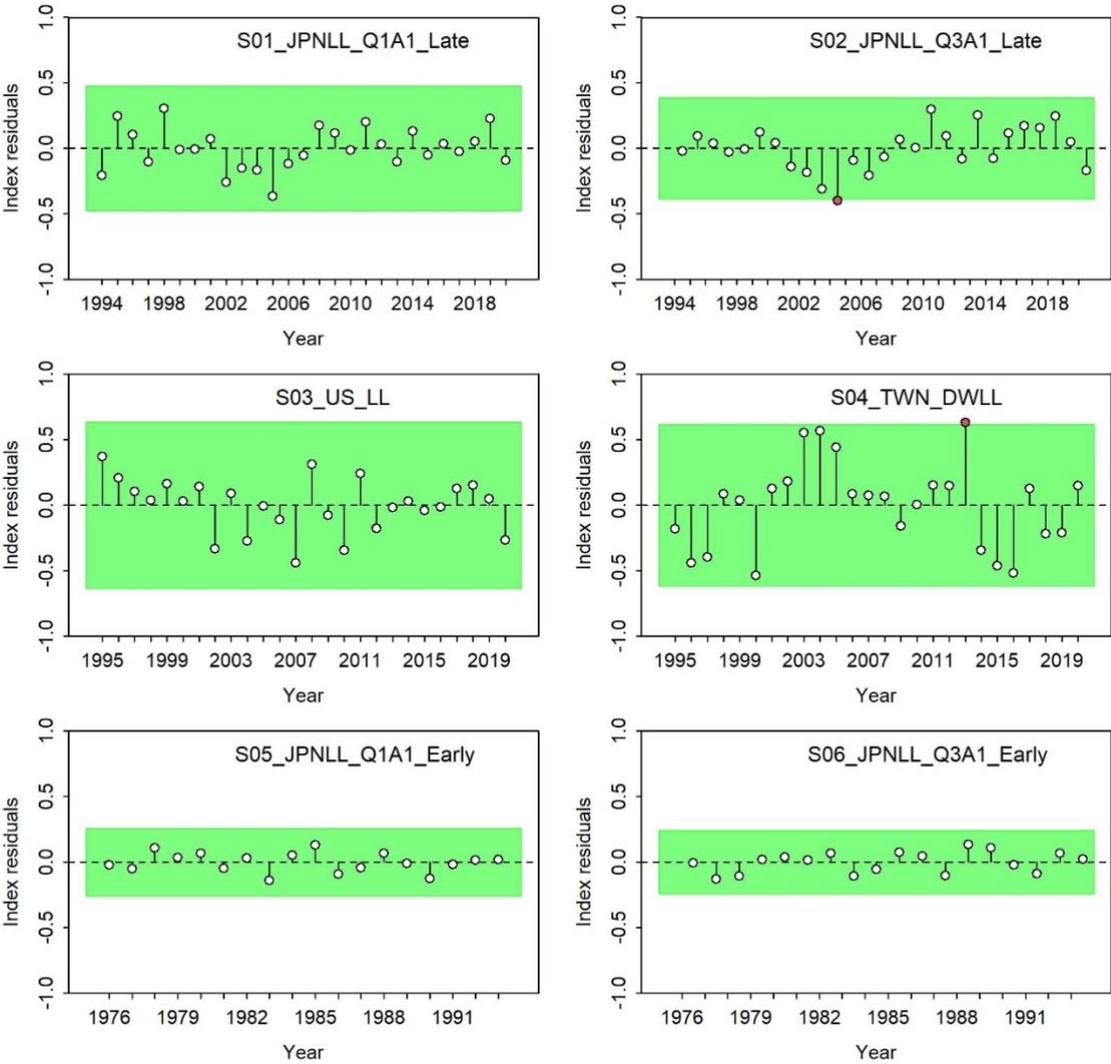


Figure 11. Results from a runs test for each CPUE index. Red indicates the index failed the test (residuals are not random), green indicates the index passed the test.

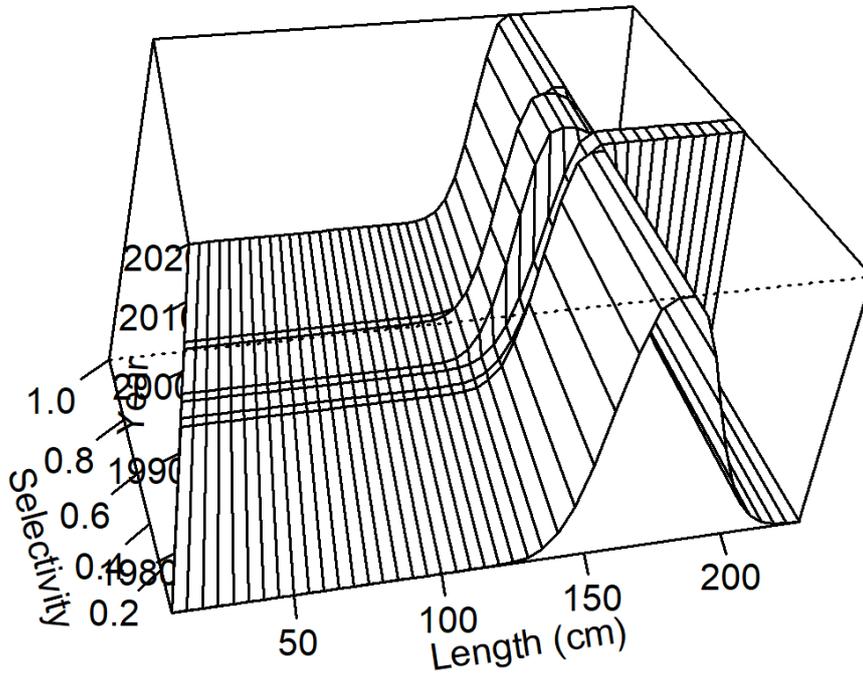


Figure 12. Time-varying selectivity estimated for F01 Japan LL Q1A1 Late.

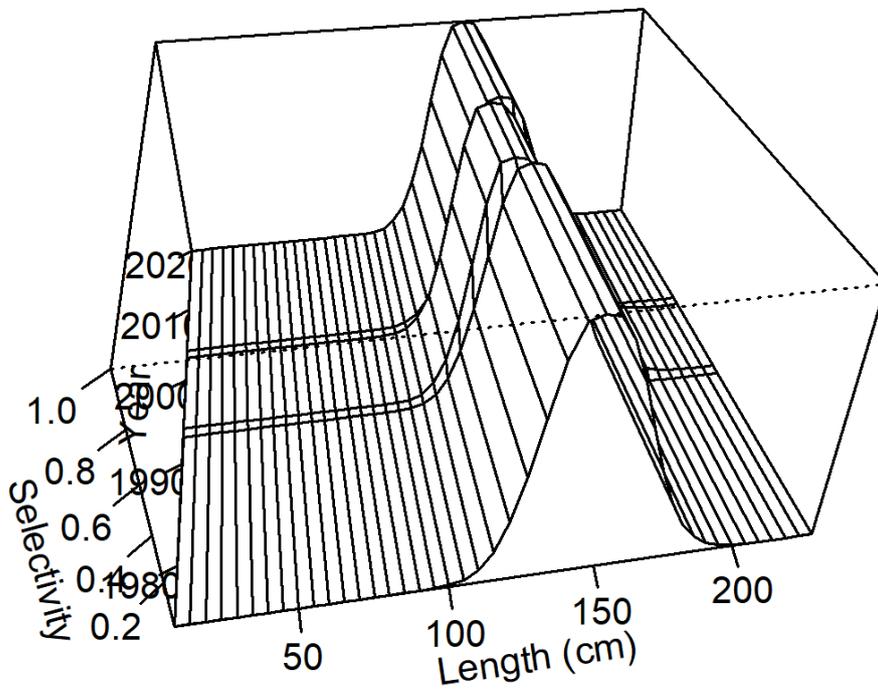


Figure 13. Time-varying selectivity estimated for F05 Japan LL Q3A1 Late.

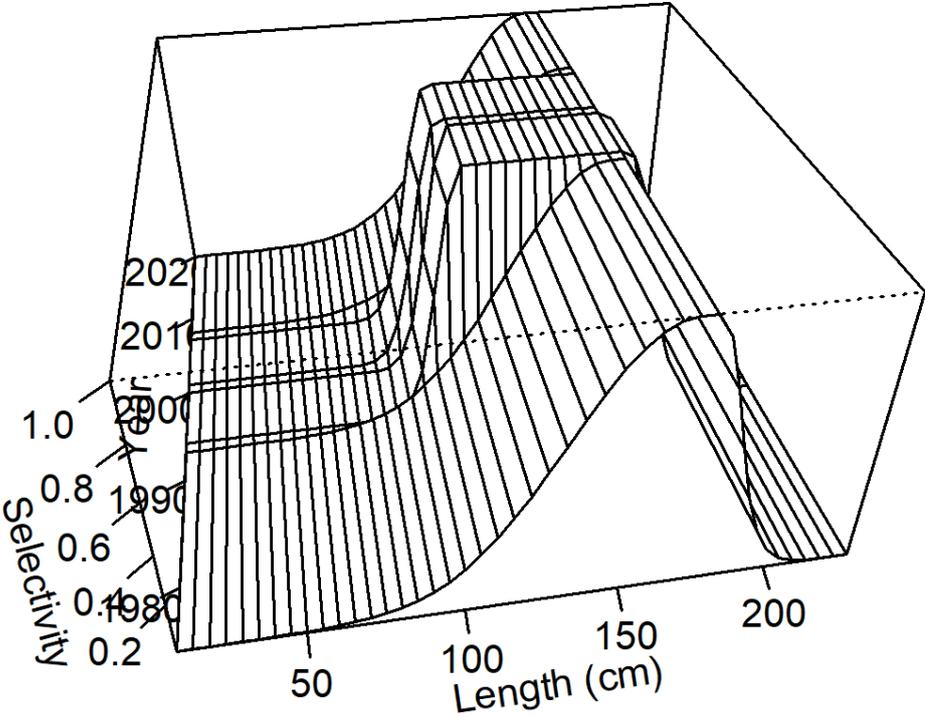


Figure 14. Time-varying selectivity estimated for F16 US Hawaii LL.

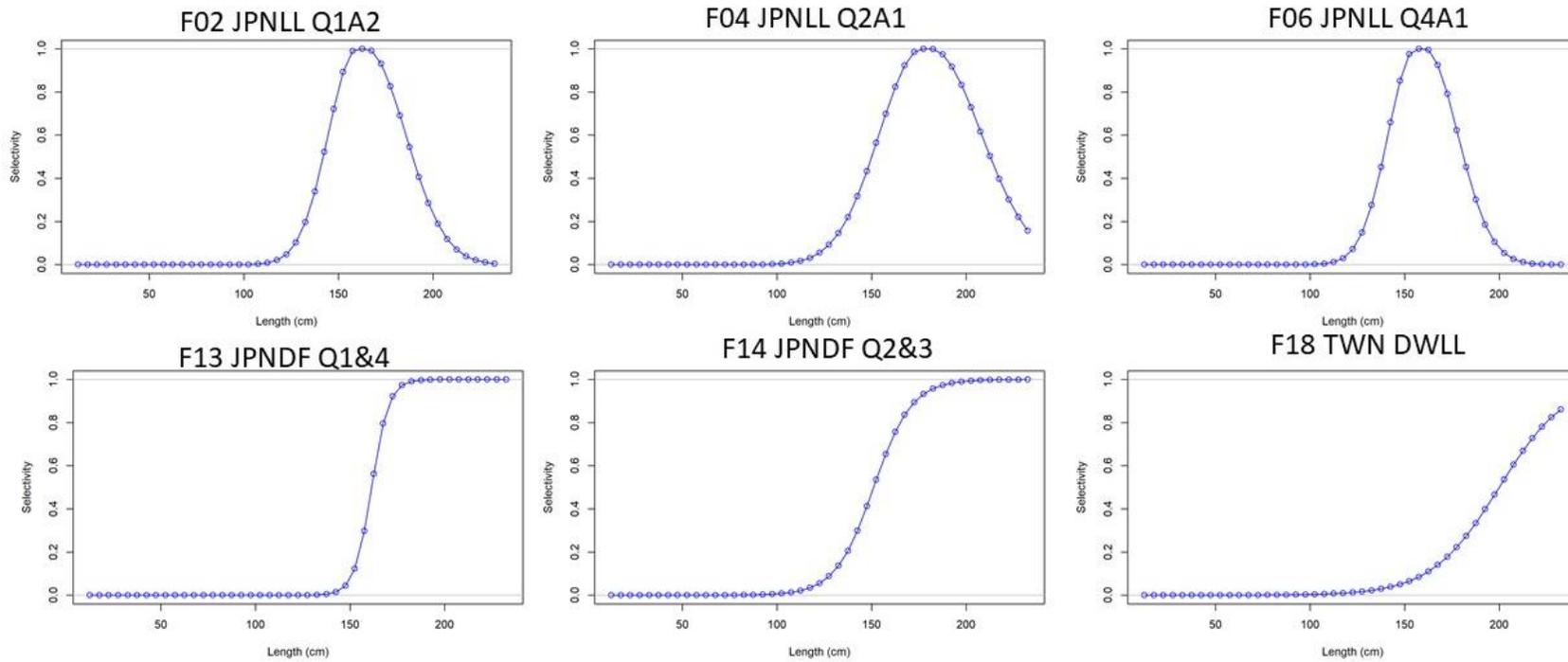


Figure15. Selectivity estimates for each of the 6 fleets without time-varying parameters. Clockwise from the top left: F02 Japan LL Q1A2, F04 Japan LL Q2A1, F06 Japan LL Q4A1, F13,Japan Driftnet Q1&4, F14 Japan Driftnet Q2&3, F18 Chinese Taipei LL.

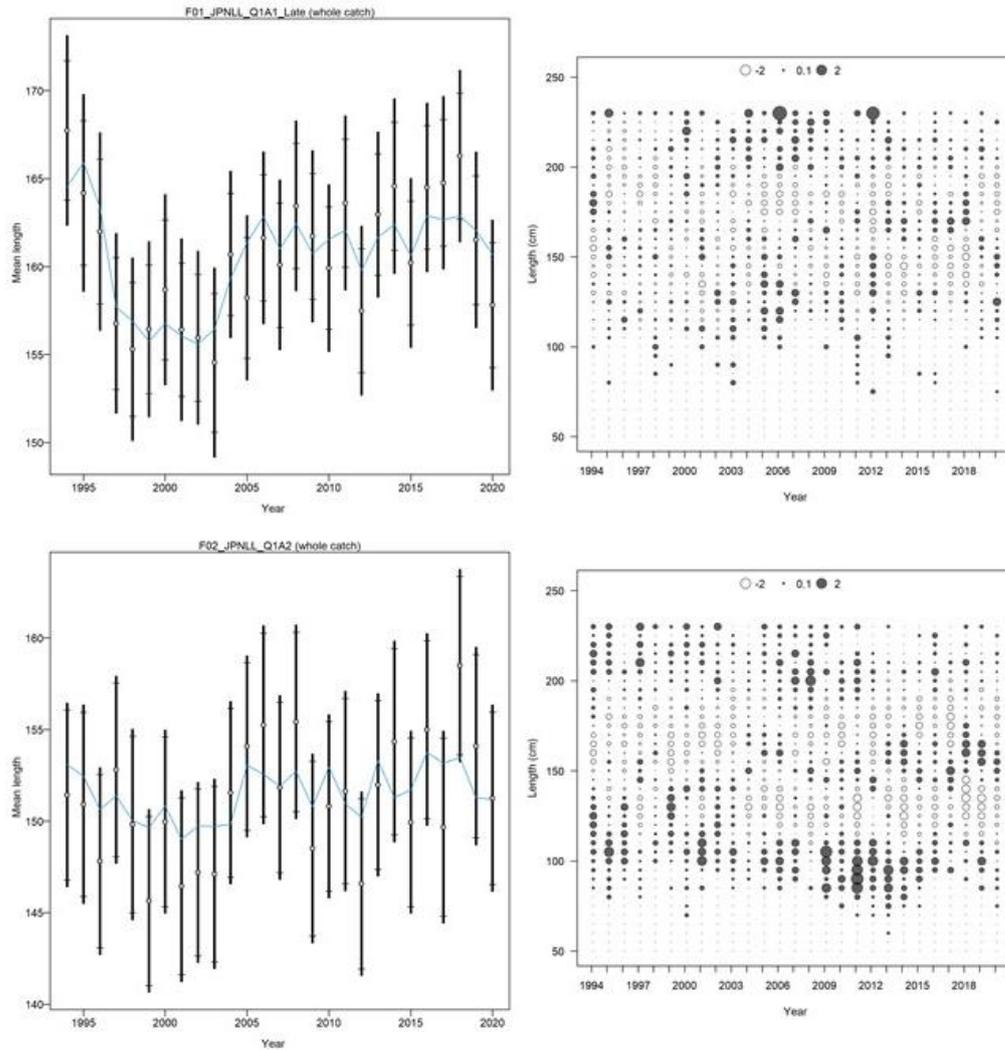


Figure 16. Fits to the annual mean length (left panels) and quarterly residuals (right panels) for Japan LL Q1A1 late (top) and Q1A2 (bottom) length composition data. The blue line indicates the estimated mean length, open dots indicate input mean length with black bars indicating the distribution of the length data with the added variance. Open circles indicate negative residuals and closed circles indicate positive residuals.

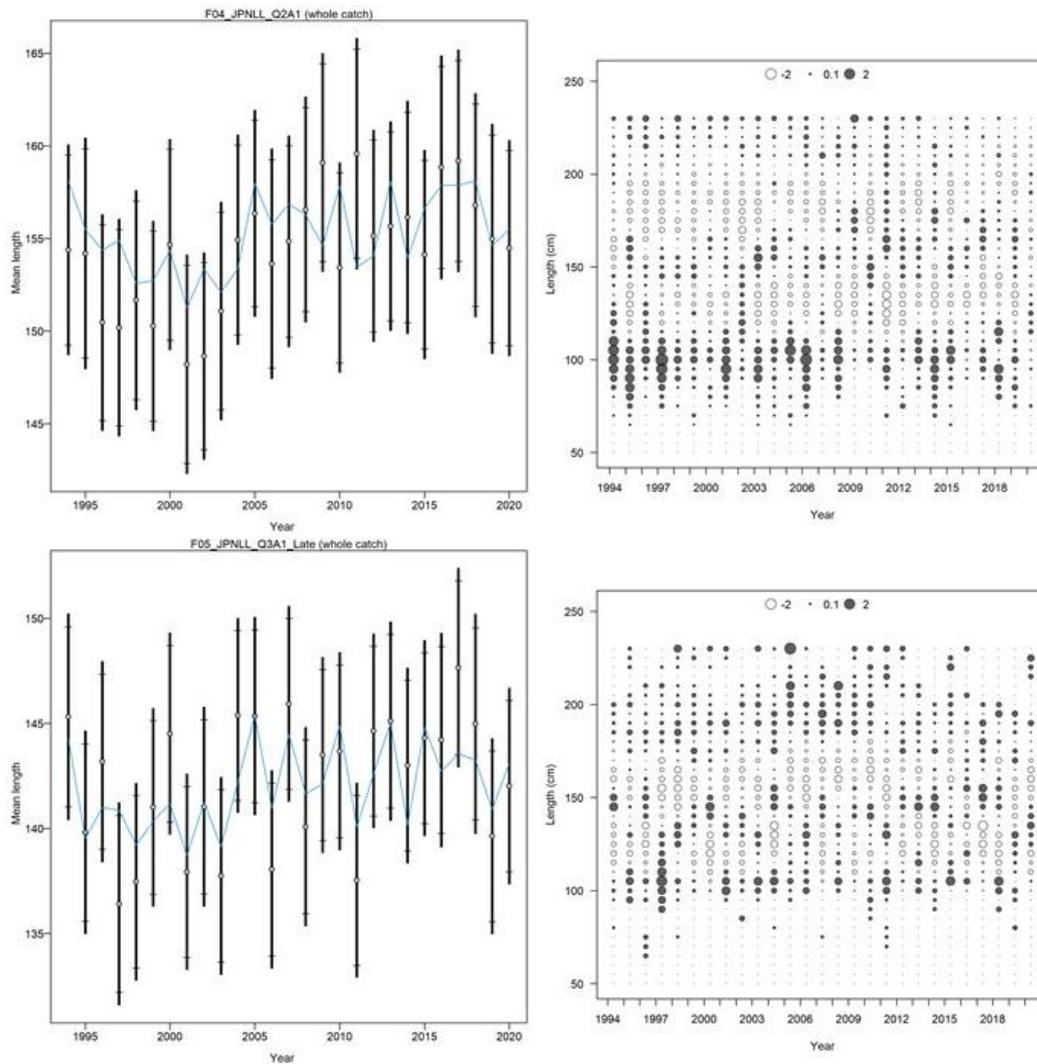


Figure 17. Fits to the annual mean length (left panels) and quarterly residuals (right panels) for Japan LL Q2A1 (top) and Q3A1 late (bottom) length composition data. The blue line indicates the estimated mean length, open dots indicate input mean length with black bars indicating the distribution of the length data with the added variance. Open circles indicate negative residuals and closed circles indicate positive residuals.

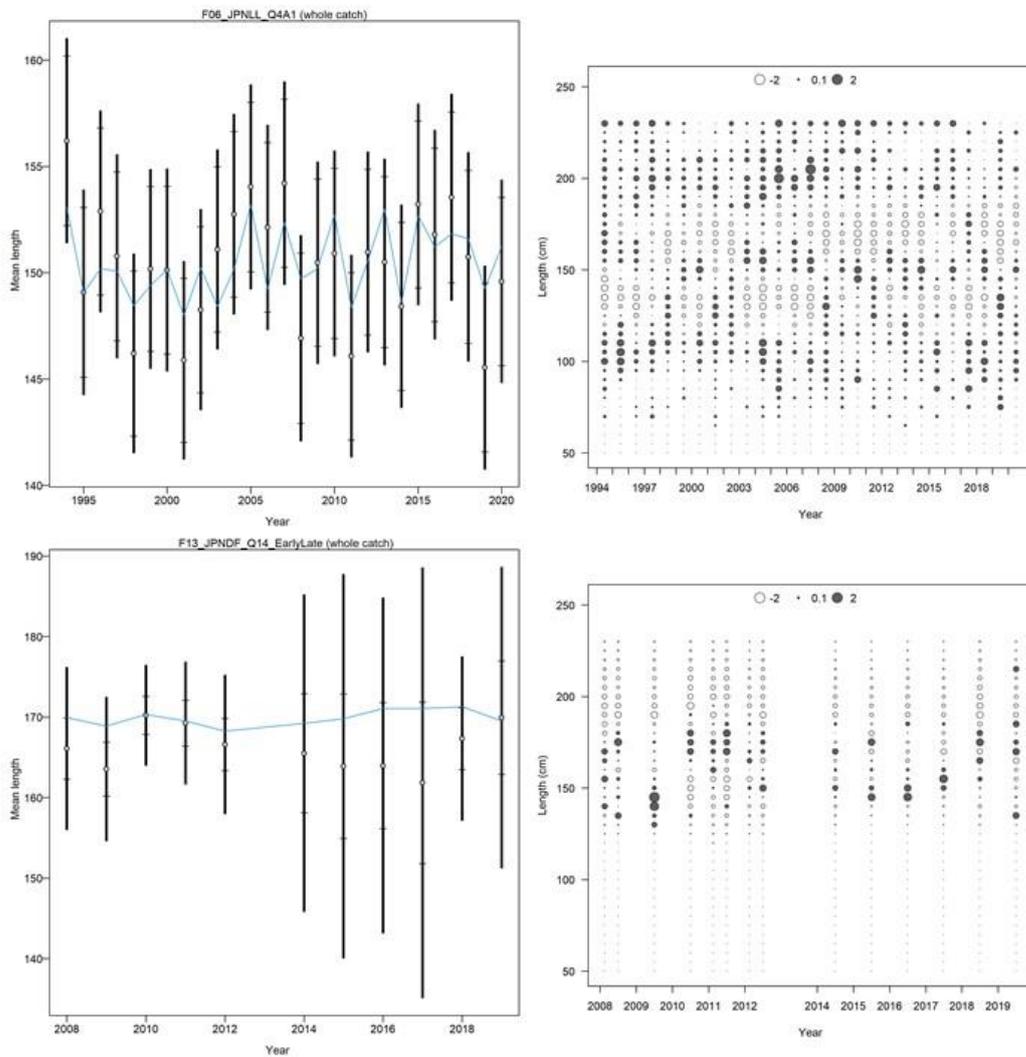


Figure 18. Fits to the annual mean length (left panels) and quarterly residuals (right panels) for Japan LL Q4A1 (top) and Japan driftnet Q1&4 (bottom) length composition data. The blue line indicates the estimated mean length, open dots indicate input mean length with black bars indicating the distribution of the length data with the added variance. Open circles indicate negative residuals and closed circles indicate positive residuals.

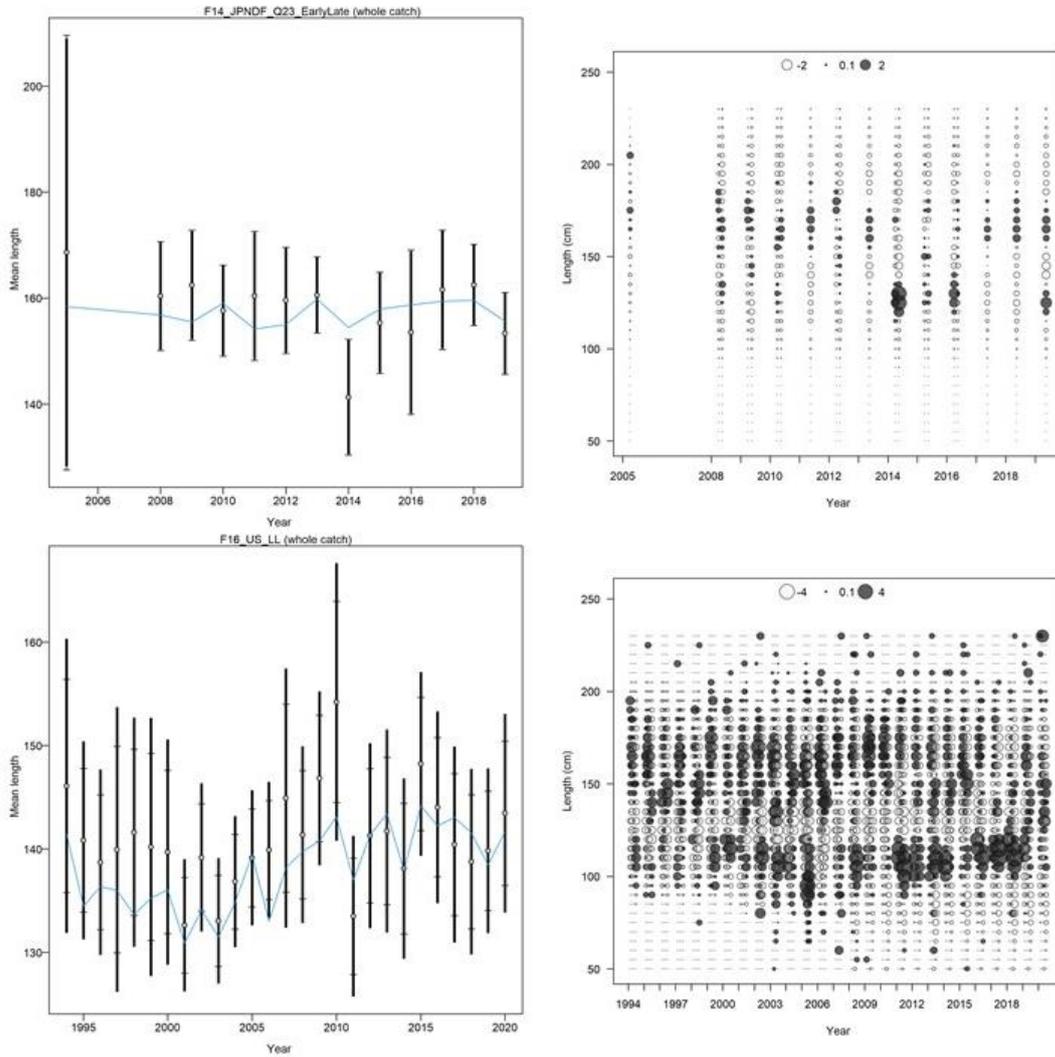


Figure 19. Fits to the annual mean length (left panels) and quarterly residuals (right panels) for Japan driftnet Q2&3 (top) and US Hawaii LL (bottom) length composition data. The blue line indicates the estimated mean length, open dots indicate input mean length with black bars indicating the distribution of the length data with the added variance. Open circles indicate negative residuals and closed circles indicate positive residuals.

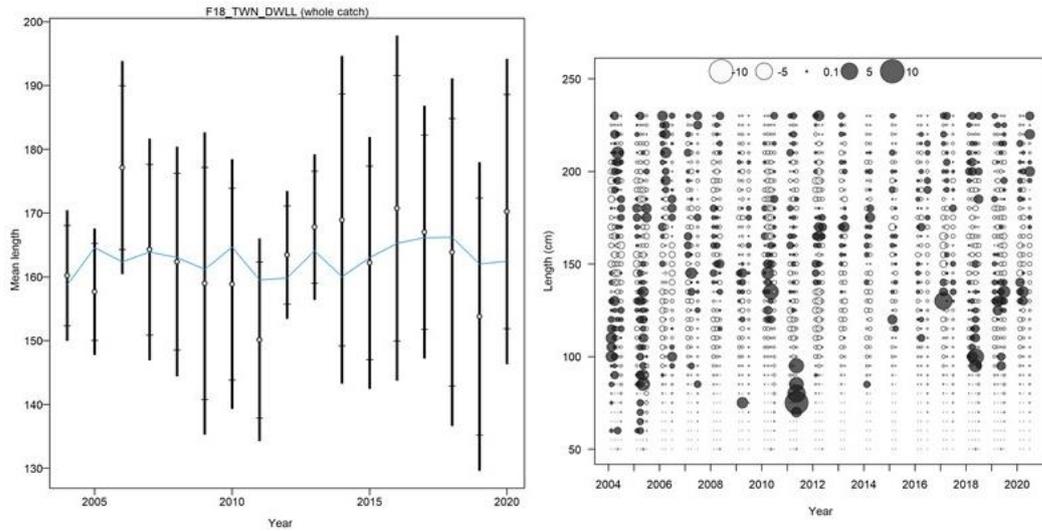


Figure 201. Fits to the annual mean weight (left panels) and quarterly residuals (right panels) for Chinese Taipei DWLL weight composition data. The blue line indicates estimated mean weight, open dots indicate input mean weight with black bars indicating the distribution of the weight data with the added variance. Open circles indicate negative residuals and closed circles indicate positive residuals.

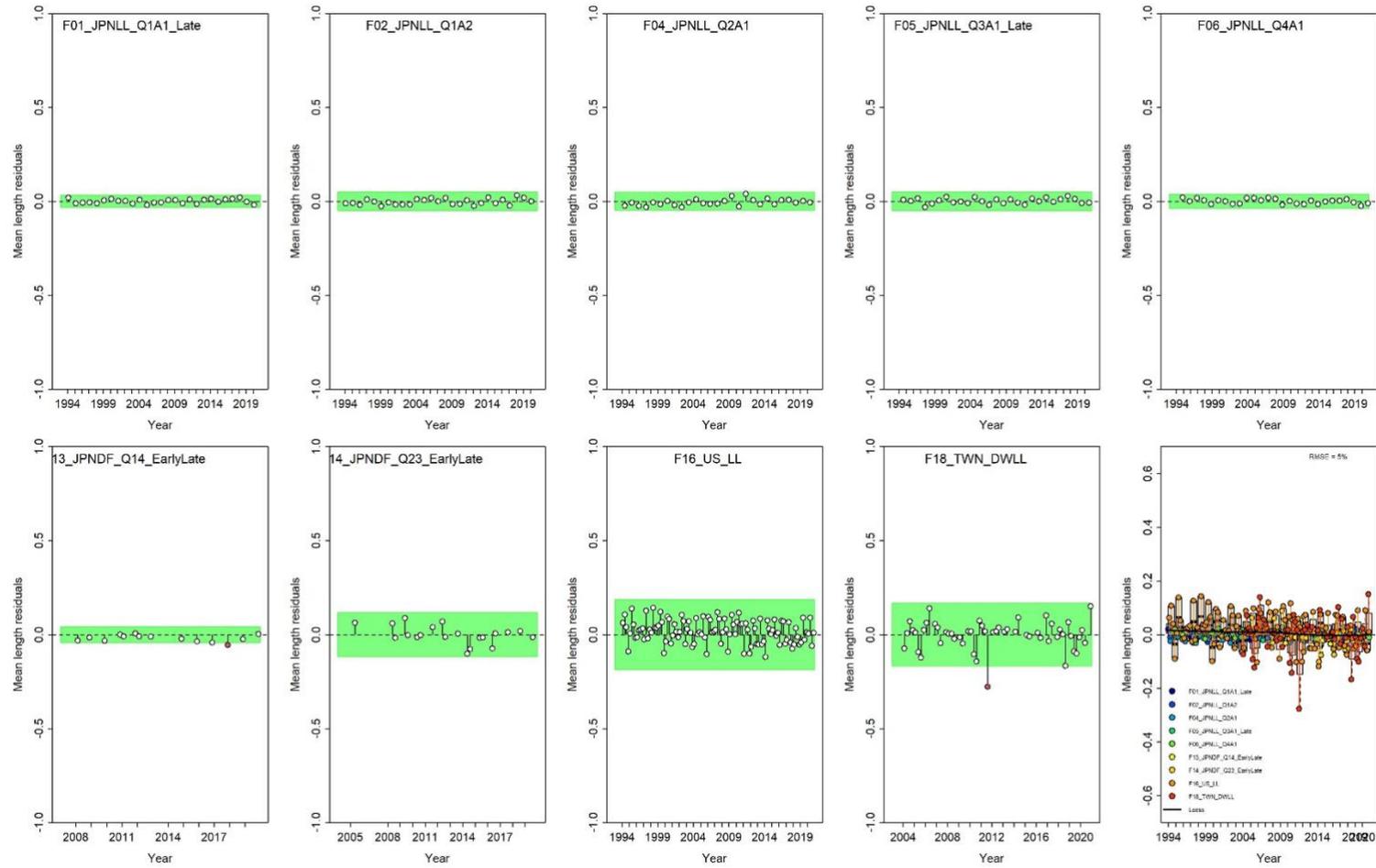


Figure 21. Results from a runs test for each size composition time series. Red indicates the data failed the test (residuals are not random), green indicates the data passed the test.

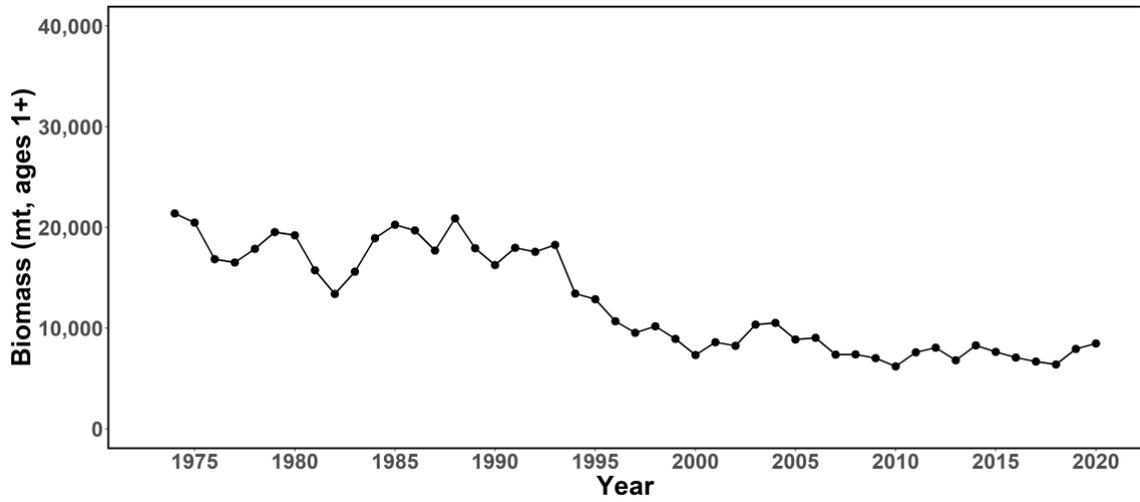


Figure 222. Estimated biomass (mt) of WCNPO striped marlin ages 1+ from the base-case model.

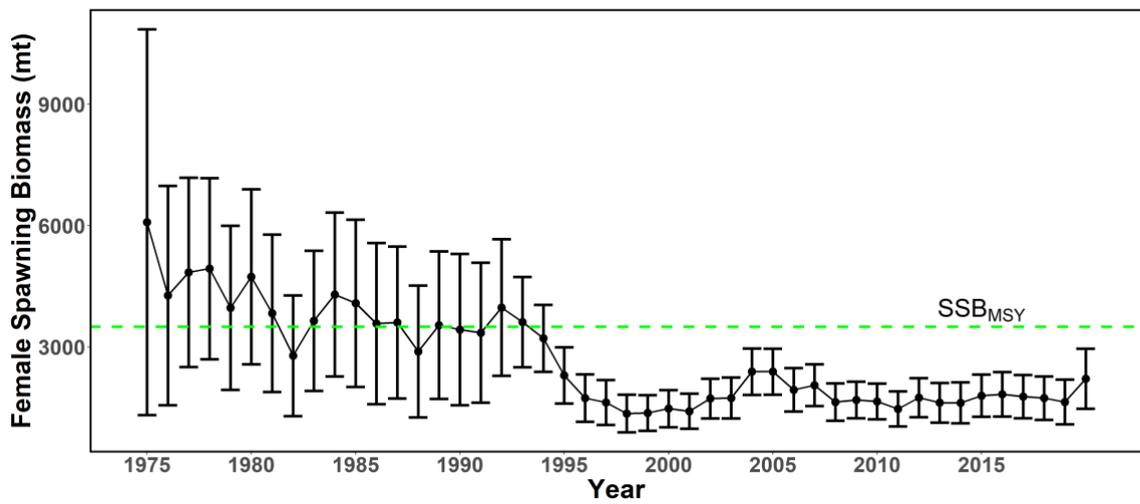


Figure 23. Estimated WCNPO striped marlin Spawning Stock Biomass (SSB) from the with 95% confidence intervals. SSB_{MSY} is indicated by the dashed green line.

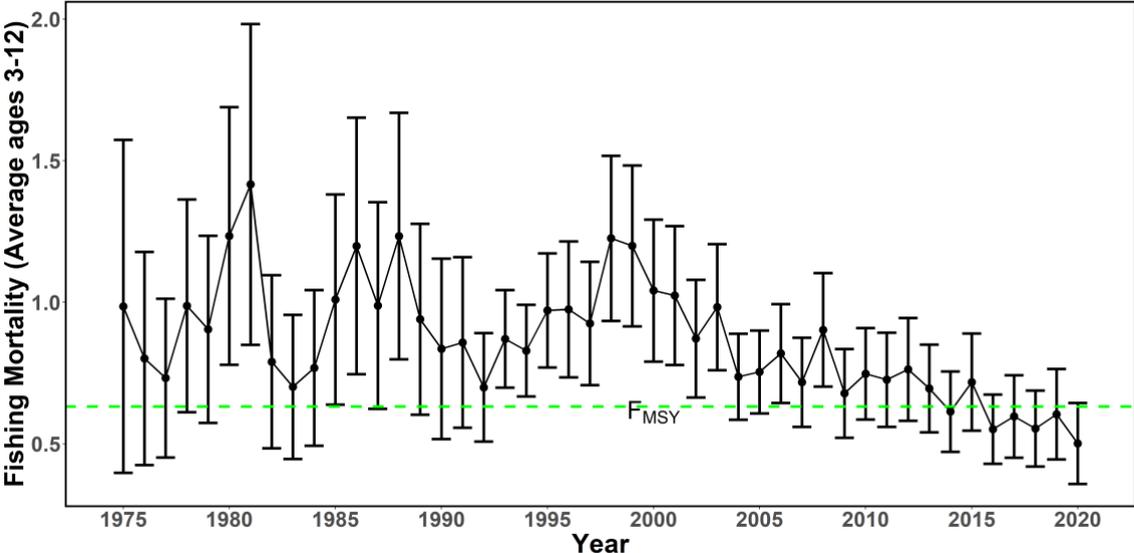


Figure 24. Estimated annual fishing mortality (ages 1-10) with 95% confidence intervals. F_{MSY} is indicated by the dashed green line.

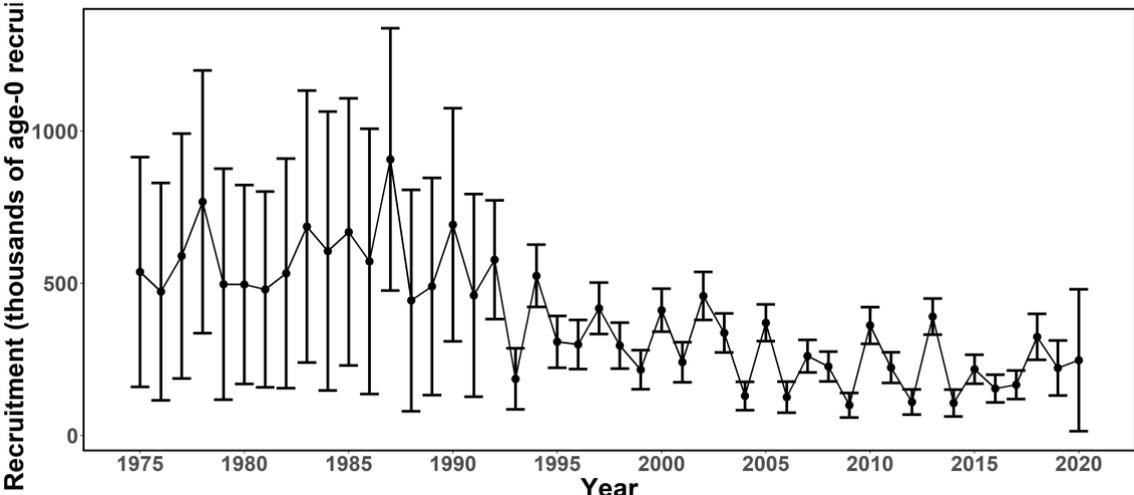


Figure 25. Estimated annual recruitment (thousands of age-0 fish) with 95% confidence intervals.

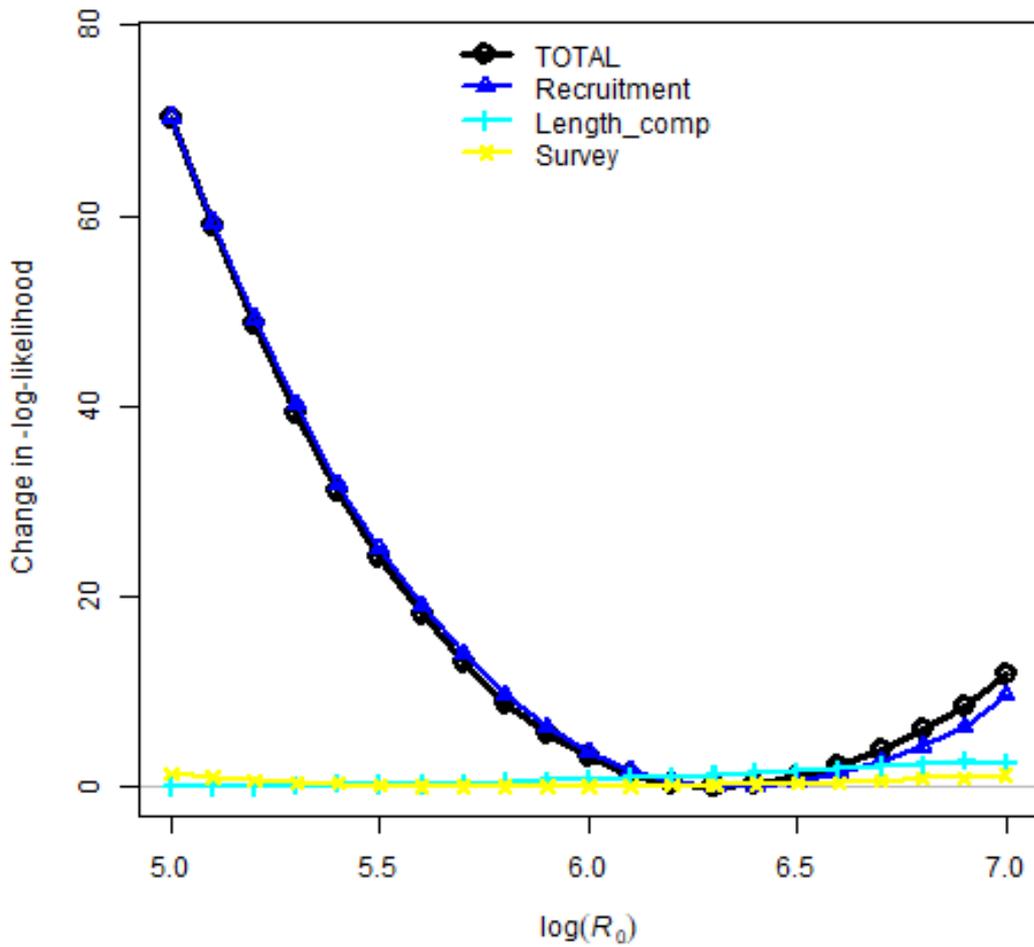


Figure 26. Likelihood profile over R_0 for the base-case model: total likelihood (black circles), recruitment (blue triangles), length composition data (light blue vertical bars), and survey/CPUE indices (yellow diamonds).

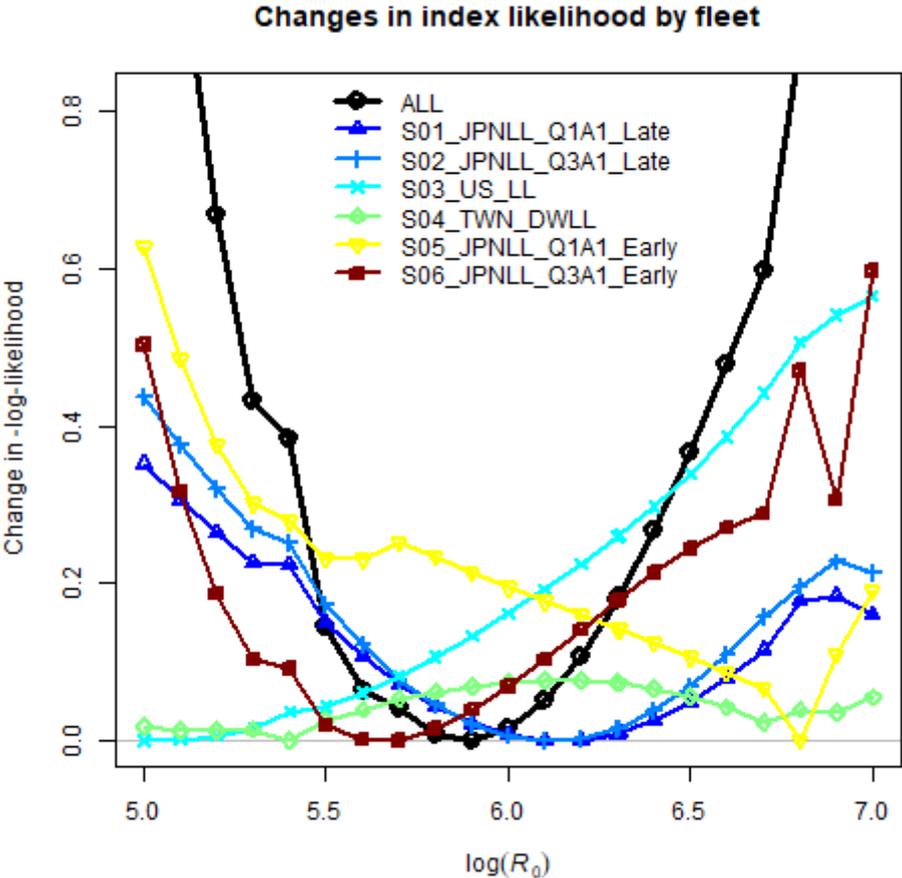


Figure 27. Likelihood profile over R₀ by CPUE index for the base-case model.

Changes in Length Composition Likelihood by fleet

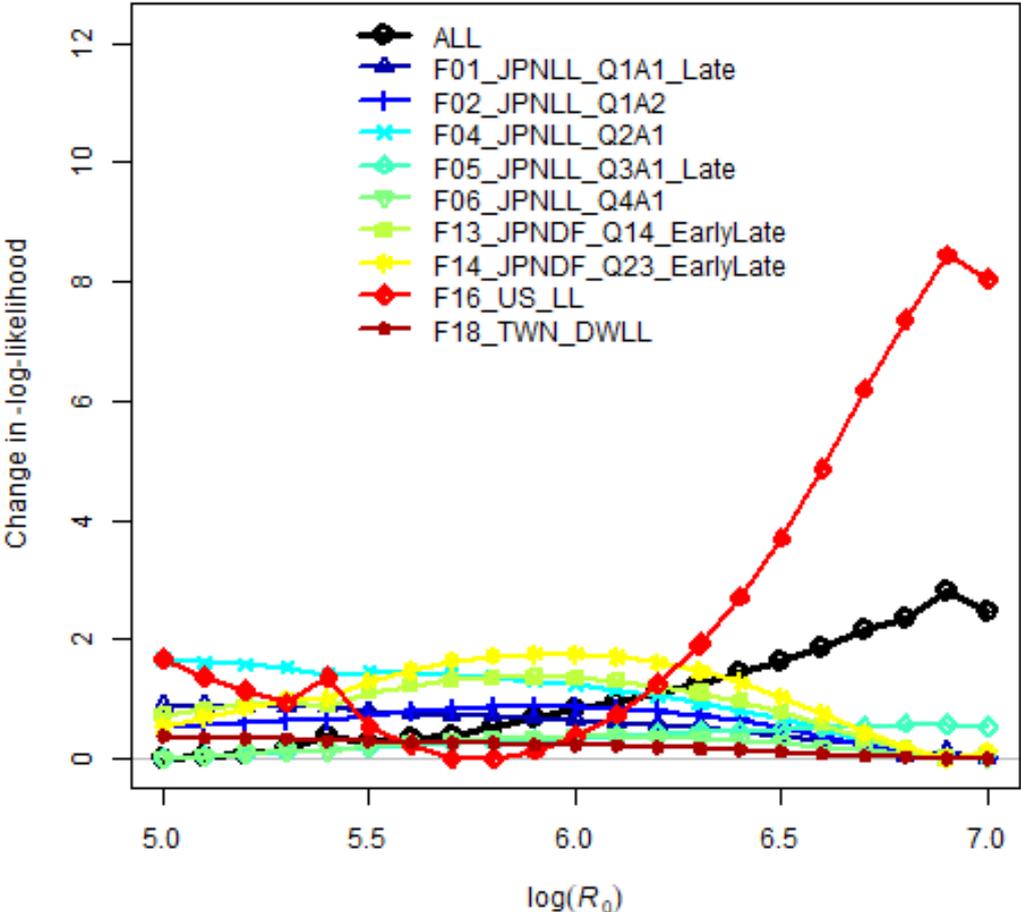


Figure 28. Likelihood profile over R0 for each length composition time series for the base-case model.

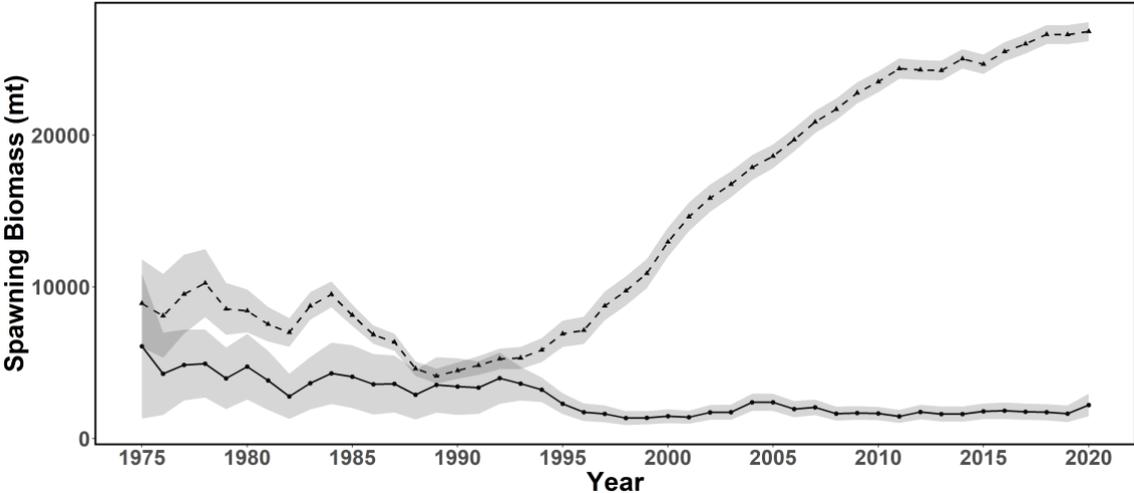


Figure 29. Spawning stock biomass trend for the ASPM model run (dashed line, triangles) and the base-case model (solid line, circles). Grey shading indicates 95% confidence intervals for each model.

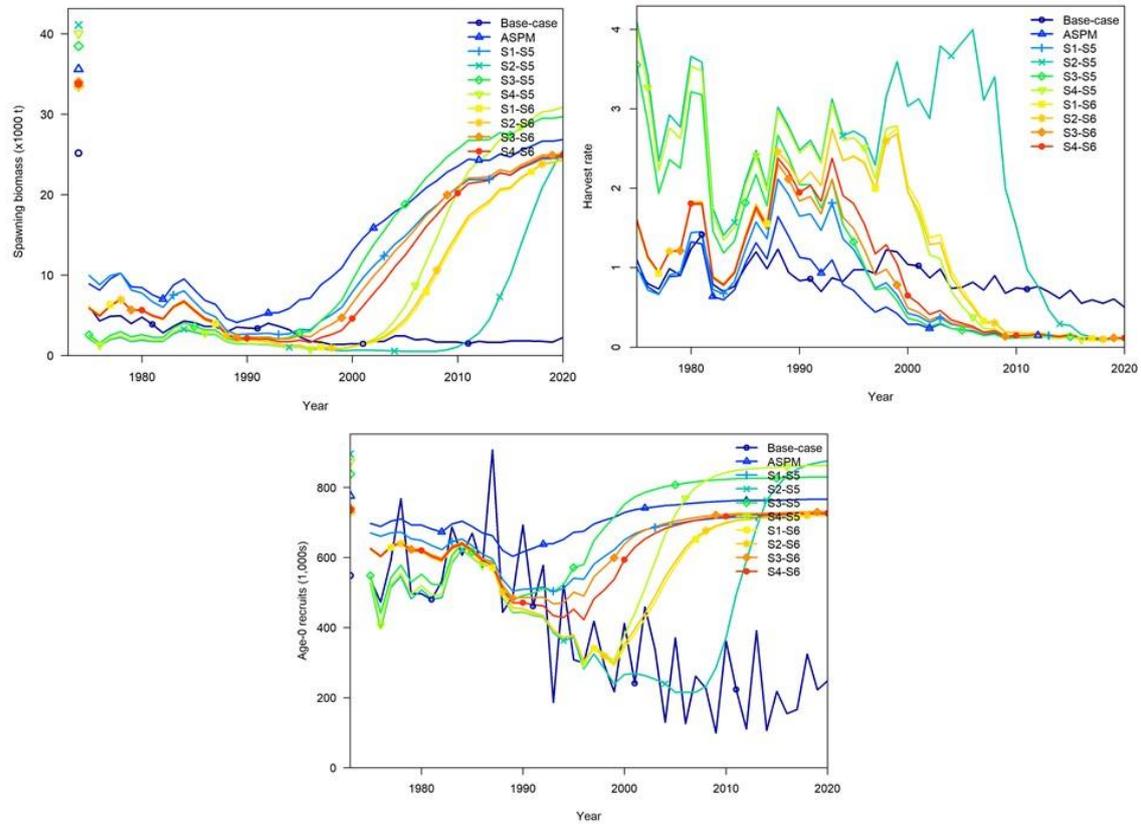


Figure 30. Comparison of a series of one-off ASPM runs using a single CPUE index for the early and late periods. Dark blue circles are the base-case model, blue triangles are the full ASPM, other runs are labels by the CPUE indices used in the model run. Top right is spawning stock biomass, top left is fishing mortality, and bottom is recruitment estimates.

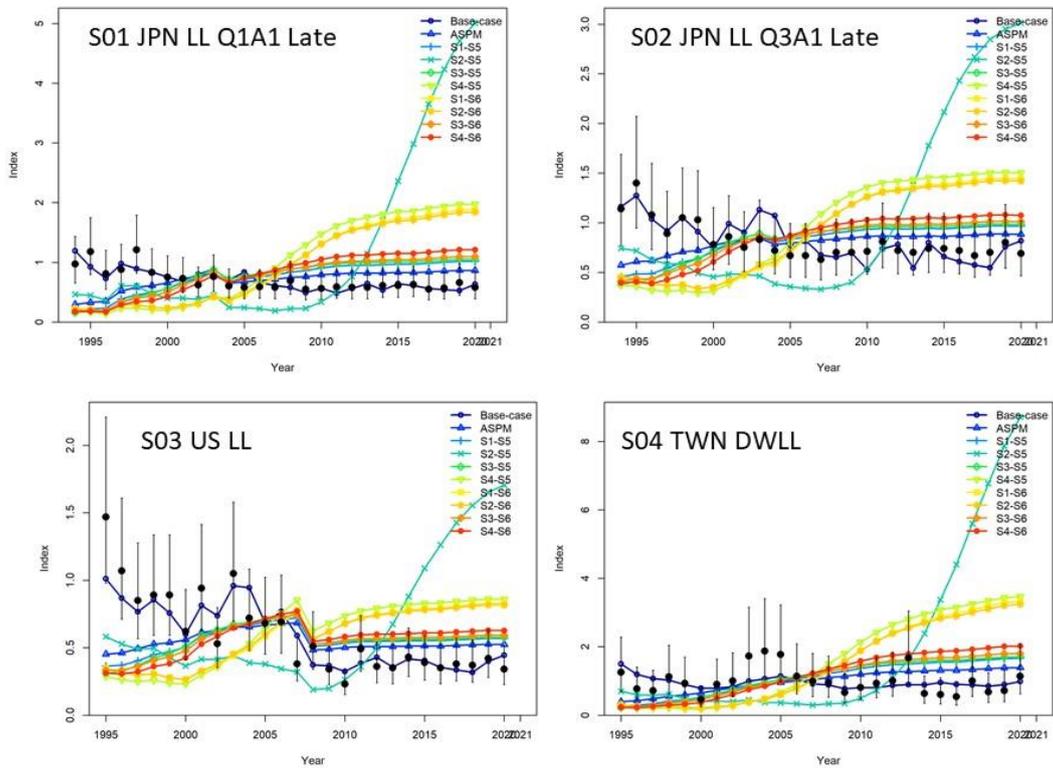


Figure 31. Comparison of a series of one-off ASPM runs using a single CPUE index for the early and late periods and their fit to each late period CPUE index. Black circles are the input CPUE index with 95% confidence intervals, dark blue circles is base-case model, blue triangles are the full ASPM, other runs are labels by the CPUE indices used in the model run.

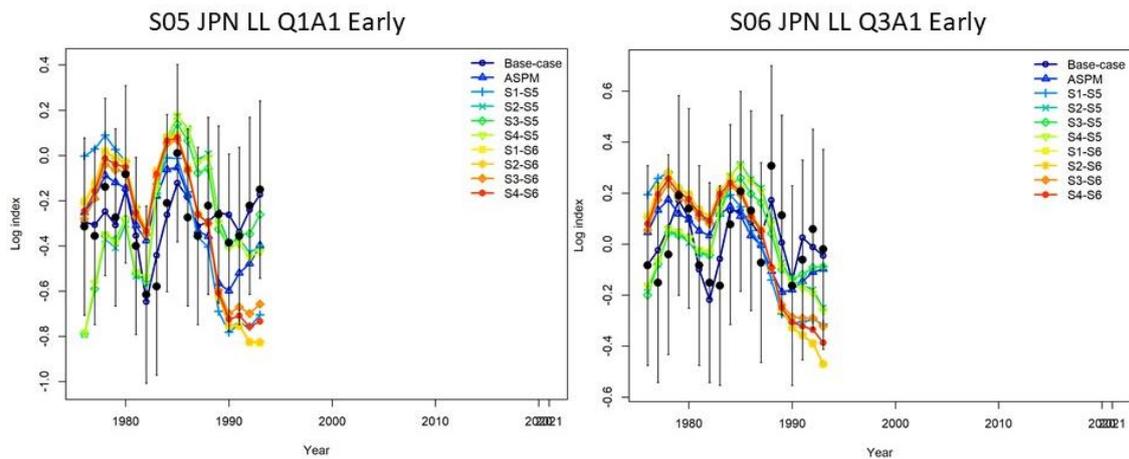


Figure 32. Comparison of a series of one-off ASPM runs using a single CPUE index for the early and late periods and their fit to each early period CPUE index. Black circles are the input CPUE index with 95% confidence intervals, dark blue circles is base-case model, blue triangles are the full ASPM, other runs are labels by the CPUE indices used in the model run.

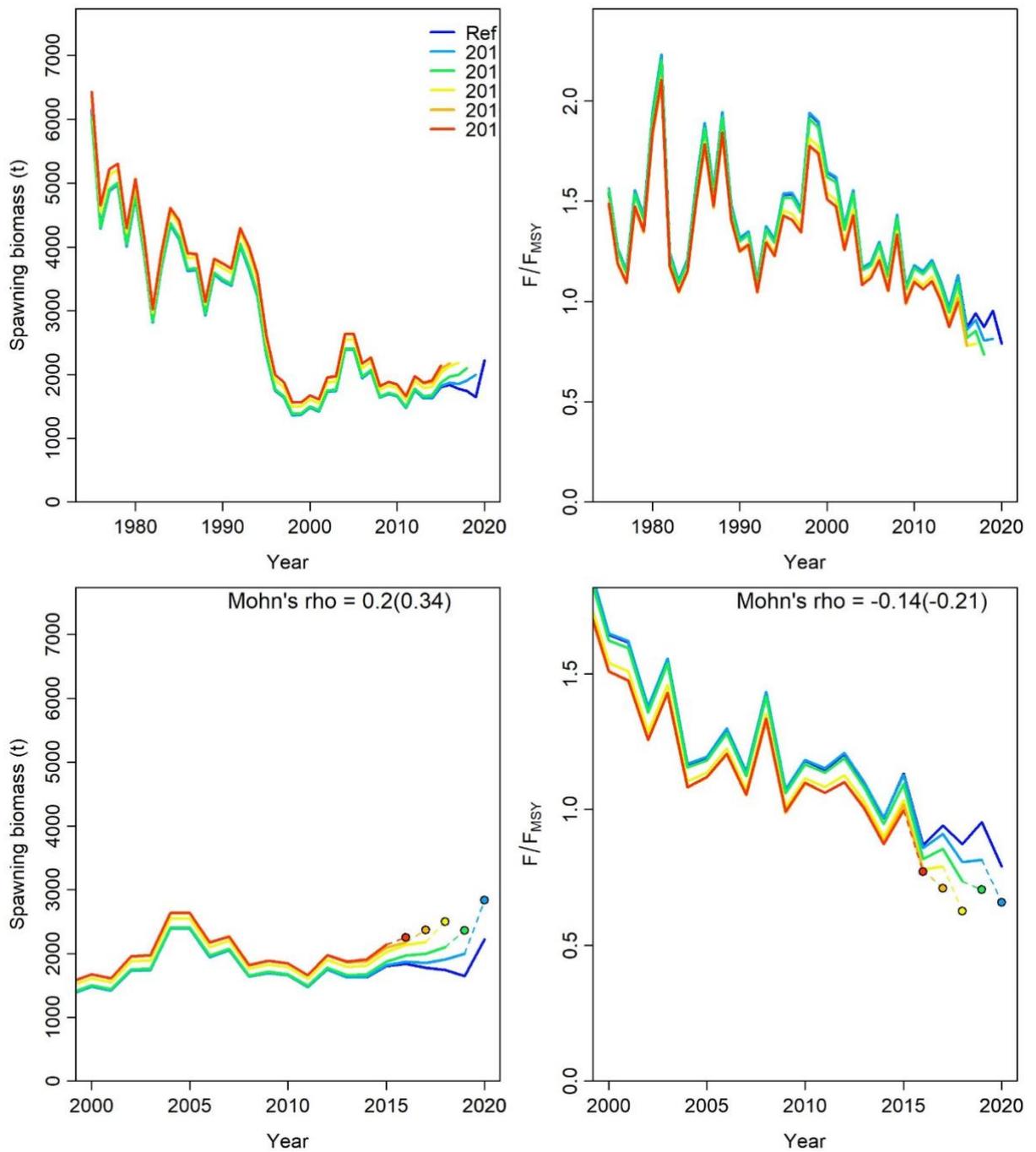


Figure 33. Retrospective analysis of spawning biomass (left) and fishing mortality (right) consisting of 5 reruns of the base case model each fitted with one more year of data removed from the base-case model. The top panels are the entire time series (1975-2020), the bottom panels are the time series since 2000 for visibility.

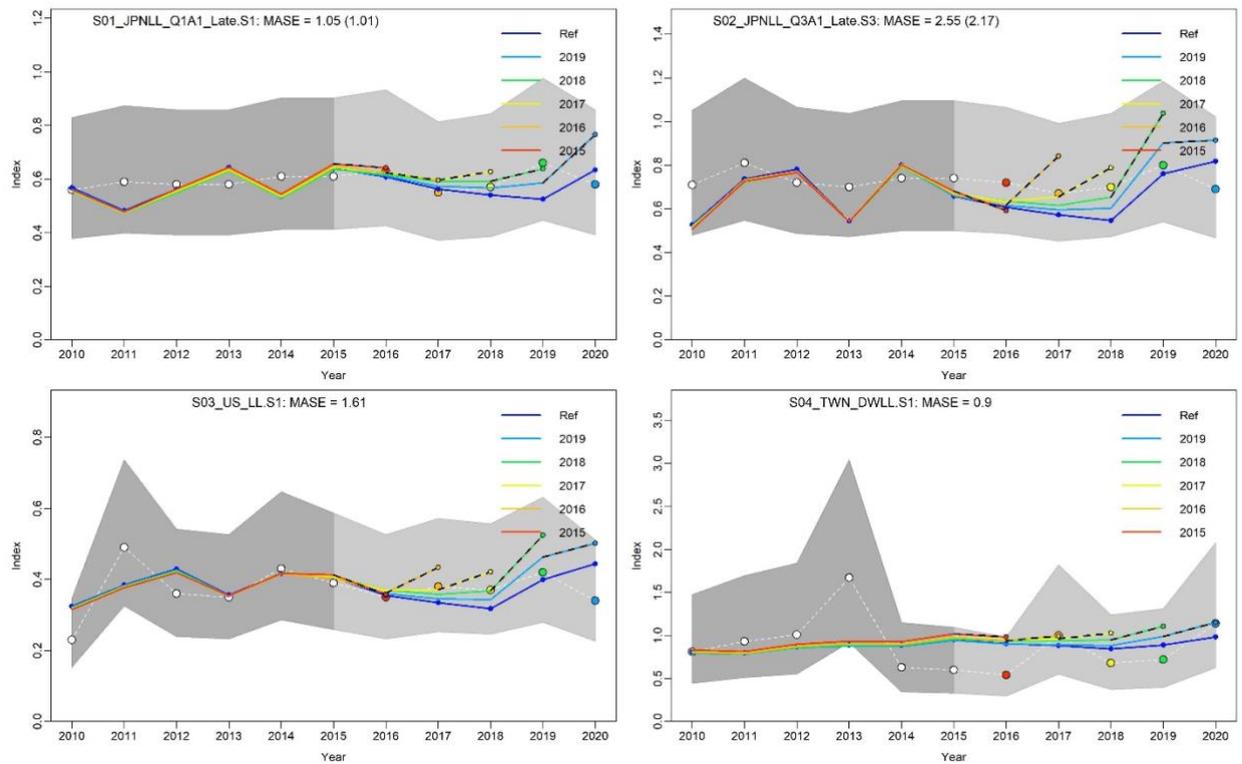


Figure 34. Hind casting cross-validation (HCxval) results for Japanese longline Q1A1 late (top right), Japanese LL Q3A1 late (top left), US Hawaii longline (bottom right), and Chinese Taipei deep water longline (bottom left) CPUE fits, showing observed (large points with dashed line), fitted (solid lines), and one-year-ahead forecast values (small terminal points) in the old growth model. The observations used for cross-validation are highlighted as color-coded solid circles with associated 95% confidence intervals (light-grey shading). The model reference year refers to the endpoint of each one-year-ahead forecast and the corresponding observation. The mean absolute scaled error (MASE) score associated with each CPUE time series is denoted in each panel.

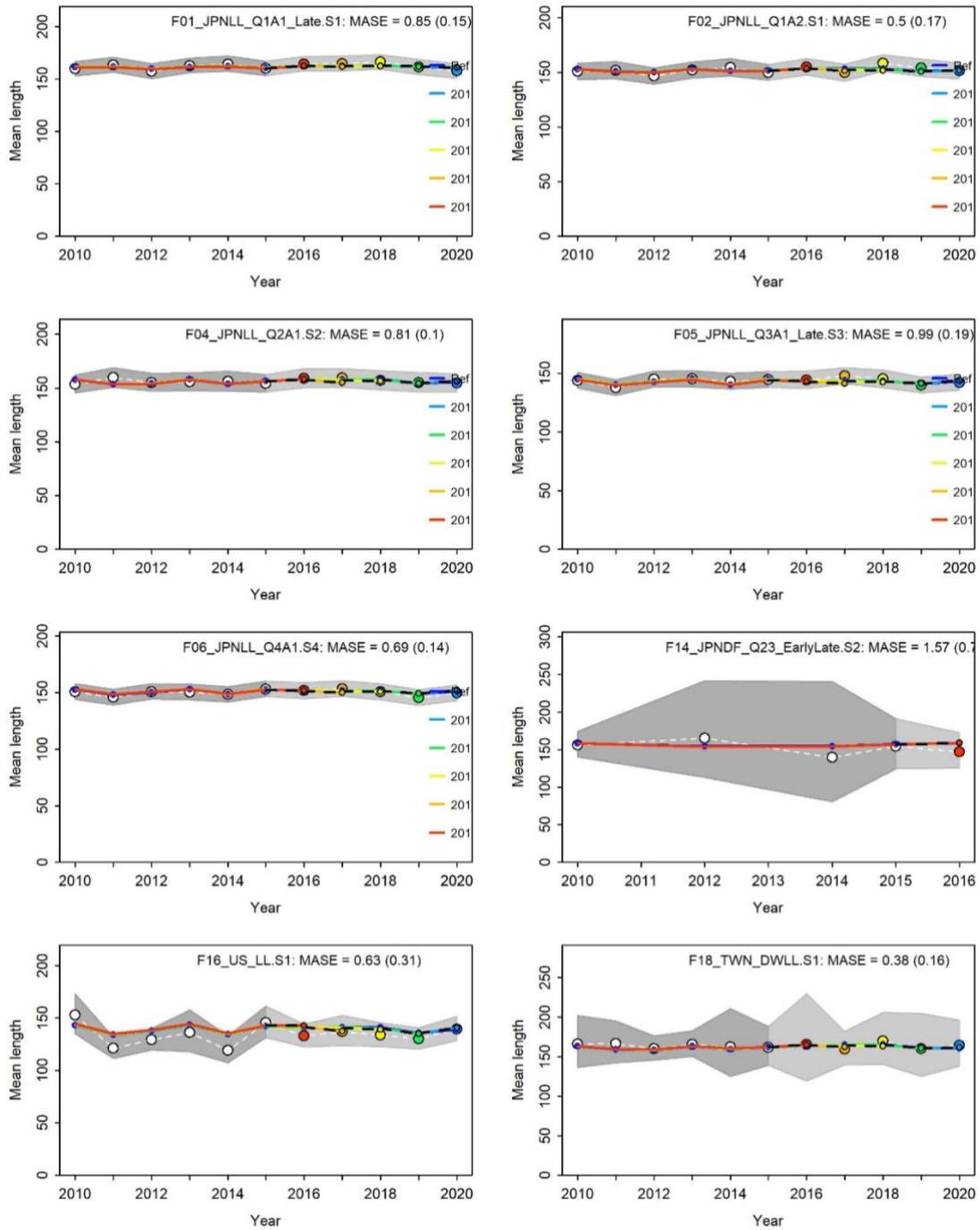


Figure 35. Hind casting cross-validation (HCxval) results for size composition mean lengths, showing observed (large points with dashed line), fitted (solid lines), and one-year-ahead forecast values (small terminal points) in the old growth model. The observations used for cross-validation are highlighted as color-coded solid circles with associated 95% confidence intervals (light-grey shading). The model reference year refers to the endpoint of each one-year-ahead forecast and the corresponding observation. The mean absolute scaled error (MASE) score associated with each size composition time series is denoted in each panel.

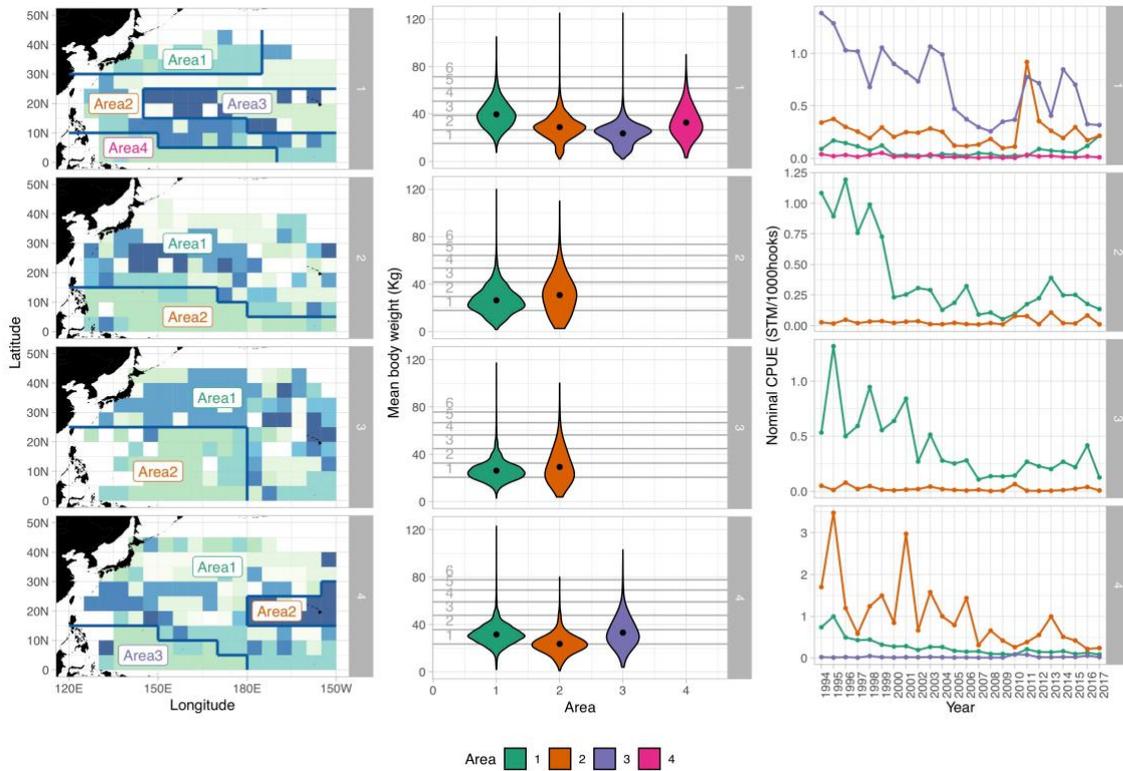


Figure 36. Figure 9 from Ijima and Kanaiwa (2019): Japanese longline fleet definition for the stock synthesis 3. Considering the results of finite mixture model analysis, we defined 11 fleets. Left panel: The area-seasonal fleet definition for WCNPO stripe marlin longline fishery. Center panel: Violin plot of mean body weight by defined area. Mean body weight is calculated by $1 \circ x \circ$ grid area. Right panel: Trends of nominal CPUE by defined area.

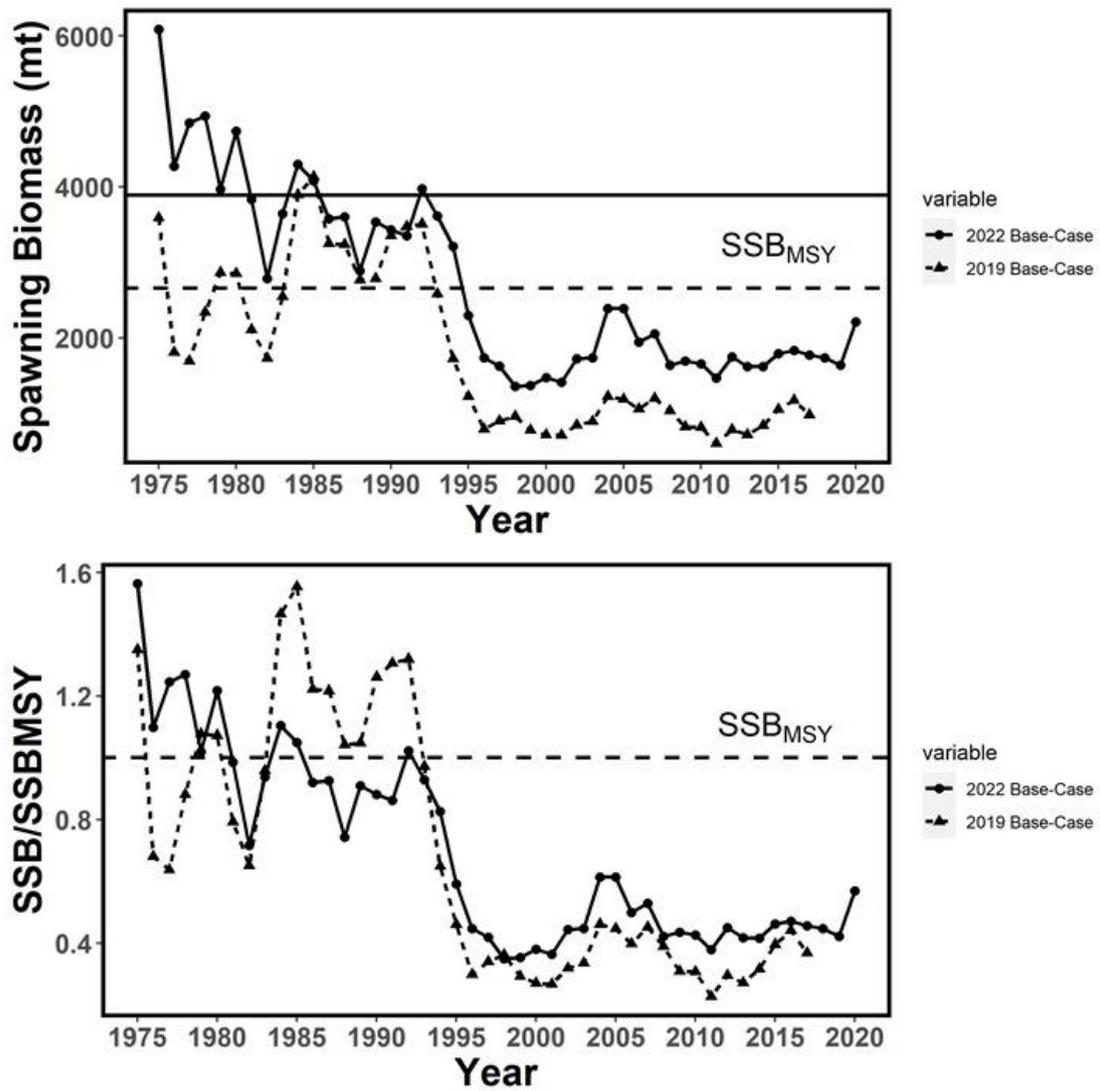


Figure 37. Comparison of spawning stock biomass (top) and SSB/SSB_{MSY} (bottom) estimated in the 2019 base-case model (dashed lines) and the 2022 base-case model (solid lines) for WCNPO striped marlin.

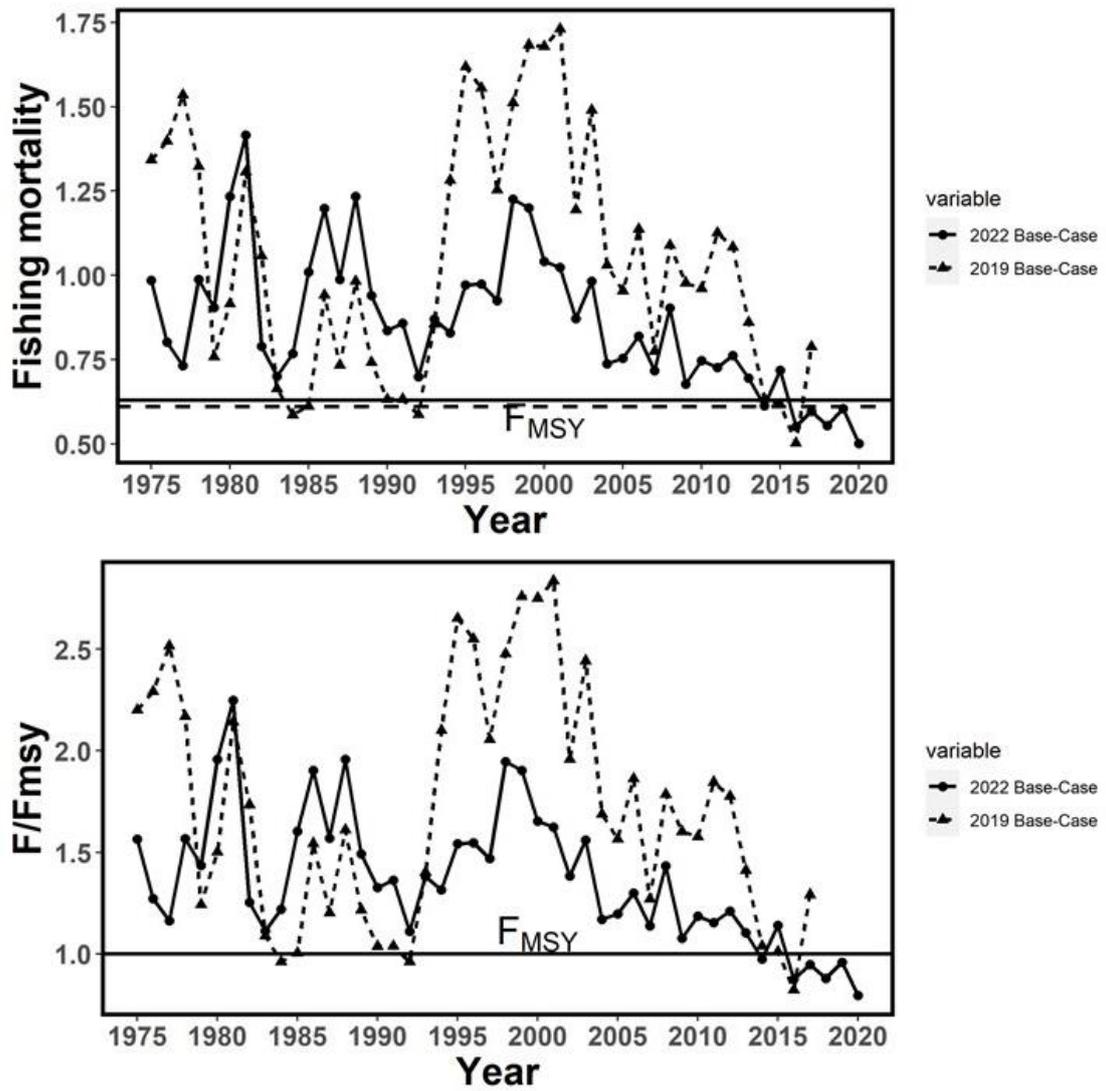


Figure 38. Comparison fishing mortality (top) and F/F_{MSY} (bottom) estimated in the 2019 base-case model (dashed lines) and the 2022 base-case model (solid lines) for WCNPO striped marlin.

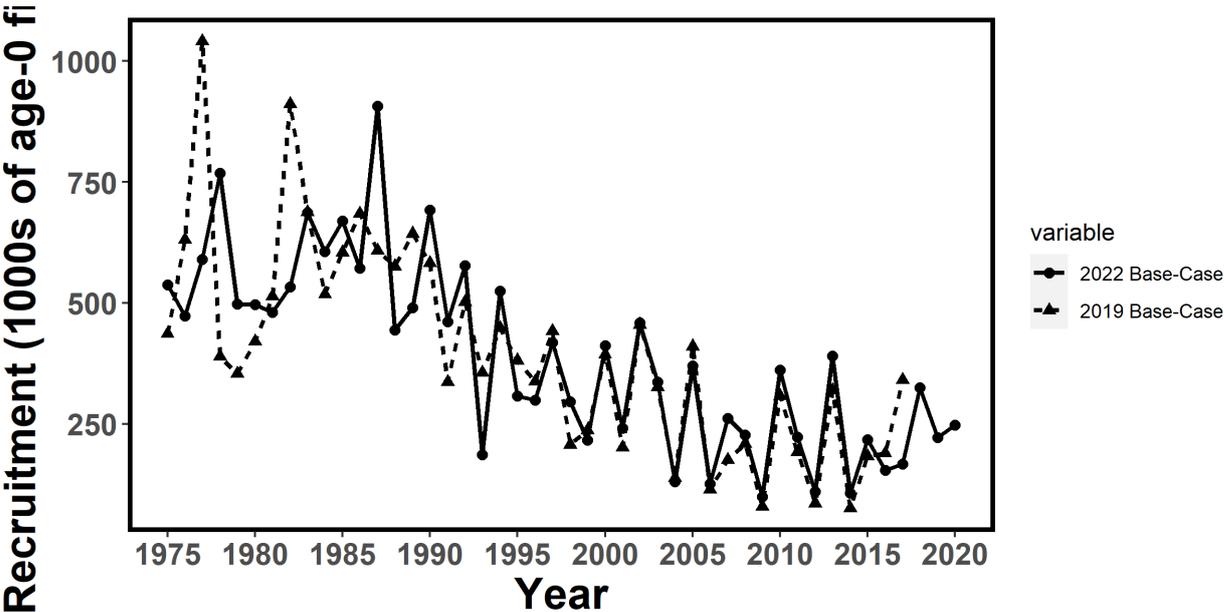


Figure 39. Comparison of estimated recruitment in the 2019 base-case model (dashed lines) and the 2022 base-case model (solid lines) for WCNPO striped marlin.

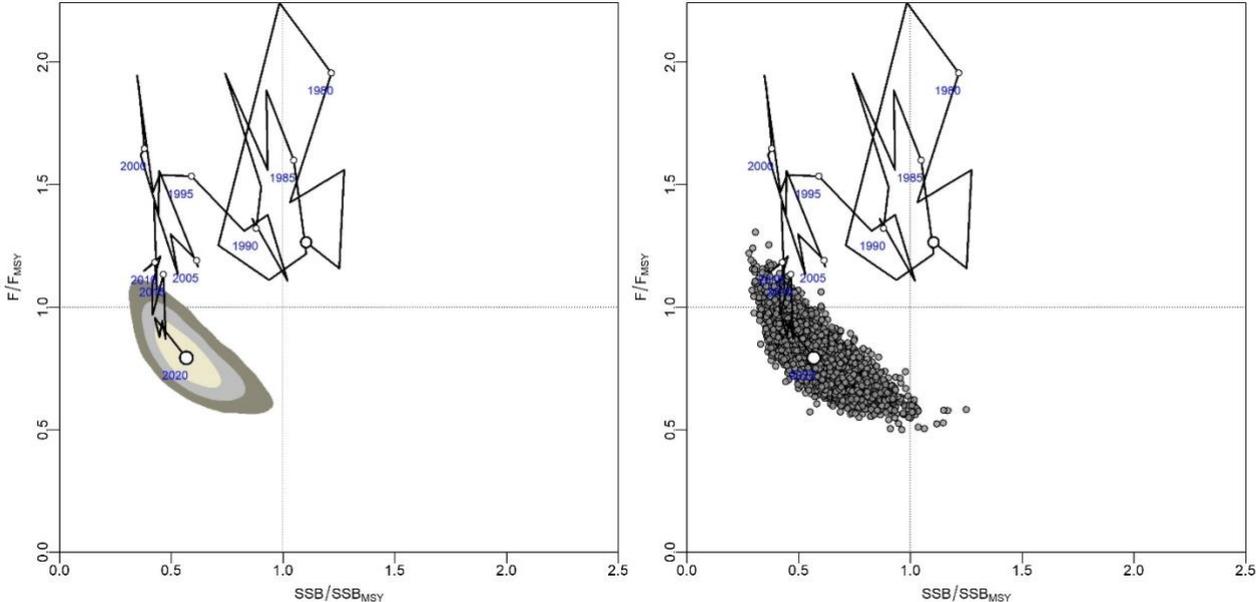


Figure 40. Kobe plot for the 2022 WCNPO striped marlin base-case model. The large white dot indicates the 2020 status, and the shaded areas indicate 50%, 80% and 95% confidence intervals (left) and dots indicate 10,000 multivariate normal draws (right) to show uncertainty around the terminal year status.